1-1-1988

Some qualifications of the alphanumeric category effect.

Christopher B. Young
University of Massachusetts Amherst

SOME QUALIFICATIONS OF THE ALPHANUMERIC CATEGORY EFFECT

A Dissertation Presented

by

CHRISTOPHER B. YOUNG

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 1988

Psychology

SOME QUALIFICATIONS OF THE ALPHANUMERIC CATEGORY EFFECT

A Dissertation Presented

by

CHRISTOPHER B. YOUNG

Approved as to style and content by:

Donald L. Fisher, Chair

Rachel K. Clifton, Member

Allen R. Hanson, Member

Alexander Pollatsek, Member

Arnold D. Well, Member

Seymour M. Berger, Department Head
Psychology

Dedicated to the Memory of Dr. C. Richard Puff

ACKNOWLEDGEMENTS

I would like to thank the chairman of my committee, Dr. Donald Fisher, for his assistance throughout this project, and at all hours of the night. I am also indebted to the other members of my committee, Dr. Alexander Pollatsek, Dr. Arnold Well, Dr. Rachel Clifton, and Dr. Allen Hanson for taking the time to make truly insightful comments and suggestions on the manuscript. I would also like to thank Chuck Clifton for the use of his extensive library and for his help with the myriad hardware and software vexations associated with this project. I think his willingness to help has been too often taken for granted.

I would certainly acknowledge John Henderson's tennis prowess, if he had taken more than a few sets from me. As it is, I can only say that he had a brutal backhand. He did, however, teach me a lot about Diner etiquette, and that conciousness is, after all, just an accident.

I thank Dolores Shank and Susan Boyce, kindred spirits, for keeping me firmly grounded in my Pennsylvanian roots; a person needs some stable moorings when all around them is in constant flux. But I believe that one can have too much of a good thing - even Pennsylvania. So, I really ought to thank Fernanda Ferreira and Sara Sereno for making me realize that there's a whole world out there - a world whose hopes, dreams, and prayers are often vastly different from my own; they kept life interesting.

Last (but by no means least), I would like to acknowledge the contribution of Dr. Randall Hansen, king of the I/O ratio, and an endurance machine. Without his careful guidance and support (especially during the lean years), I might have been writing this last year. Then again, depending on how you look at it, I might not have written it at all.

ABSTRACT

SOME QUALIFICATIONS OF THE ALPHANUMERIC CATEGORY EFFECT

SEPTEMBER 1988

CHRISTOPHER B. YOUNG, B.A., FRANKLIN AND MARSHALL COLLEGE

M.S., UNIVERSITY OF MASSACHUSETTS

Ph.D., UNIVERSITY OF MASSACHUSETTS

Directed by: Professor Donald L. Fisher

Whether the alphanumeric category effect in visual search tasks is due to physical or "conceptual" differences between target and distractor categories has been a matter of long-standing debate. Typically, subjects can search for digits in a letter background or vice-versa (between-category condition) more efficiently than for targets in a background of same-category distractors (within-category condition). Some recent work by Krueger (1984) indicates that the effect is mediated entirely by physical feature differences between the digit and letter categories. In the present study, subjects were presented with brief (175 ms) visual displays of two, four, or six alphanumeric characters. Subjects then made a speeded (button-press) response indicating the presence or absence in the display of items in a search (memory) set defined prior to the onset of the display.

In Experiment 1, parallel search functions (i.e., functions exhibiting very little increase in response time with increases in display size) were observed with two memory set sizes (one and four)

in between-category conditions, but not in within-category conditions. In Experiment 2, the effect was obtained even when target-background featural differences were controlled (in a manner similar to Krueger, 1984). Based on a significant difference (in RT means, but not slopes) between the two between-category memory set size conditions, it was argued that the effect is due to physical features when the memory set consists of a single target and category membership when there are multiple targets in the memory set. This conclusion was confirmed by catch trial data from Experiment 3. When the memory set consisted of more than one item, nine of fourteen subjects incorrectly responded "present" to a same-category (but featurally-discrepant from the memory set) foil. When the memory set consisted of a single item, none of fourteen subjects incorrectly responded "present" to this same foil. Alternative explanations of the results and some methodological considerations were discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

One of the most fundamental issues in the literature on human attentional processes is the early/late selection dichotomy. Proponents of "early" selection theories maintain that subjects select from competing stimulus inputs (or "channels") based on simple physical properties of the stimulus, e.g., pitch, location, or color (see, for example, Treisman, 1964). This implies that a stimulus (e.g., a letter) can be selected from among other inputs before the inputs make contact with their long-term memory representations if appropriate physical cues are present (e.g., an "A" in a background of "C"'s). Conversely, "late" selection theorists maintain that stimuli are selected for after they "make contact" with their long-term memory representations (see, for example, Deutsch and Deutsch, 1963).

The ability of subjects to search "in parallel" for digits among letters (or letters among digits) and the relative inability of subjects to search in parallel for targets in a background of same-category distractors (e.g., letters in letters) has been ascribed to the existence of a "semantic" category difference between digits and letters (Egeth, Jonides, and Wall, 1972). As such, the "category effect" in visual search has until recently been offered as a shining example of late selection at work. Unfortunately, the picture is no longer so clear. Recent studies (e.g., Krueger, 1984) have suggested that the effect is due entirely to "uncontrolled physical feature differences" between the target and distractor sets. If this is the case, then selection in these tasks is based on simple physical features and occurs (relatively) early in the course of information processing. To the extent that we can arbitrate between these two explanations (semantic category vs. physical features) of the effect, we gain insight into the locus of selection question.

Review of previous findings

There is now a substantial body of data demonstrating a "category effect" in visual search tasks. In the typical visual search task discussed here, subjects are asked to search for one or more "targets" (usually alphanumeric characters) which may or may not be present in a background of "distractors" (also alphanumeric characters). The subject is first presented with the "memory set" [the target(s) for that trial] for some interval (either controlled by the experimenter or self-paced). After presentation of the memory set, the subject is presented with a brief (typically 150-200 ms) display which may or may not contain one of the members of the memory set. The subject's task is to respond "present" (there was a target in the display) or "absent" (no target in the display) as quickly and as accurately as possible. By varying the target/distractor relationship and looking at the effect on the response time, one gains insight into the perceptual processes involved in the task. When subjects are asked to search for digits in a background of letters or letters in a background of digits (between-category condition), their performance is much better than when the target and background items are drawn from the same category (within-category, e.g., both letters).

Perhaps the best-known and most often-cited demonstration of the category effect is a study by Jonides and Gleitman (1972). Jonides and Gleitman had subjects search for the letters "A", "Z", and "O" (oh) or the numbers "4", "2", and "0" (zero) in a background of digits or letters. Either a single target was present on a trial or the target was absent, and the subject's task was to judge "present" or "absent" as quickly as possible. One of the three targets was specified verbally prior to each trial. Jonides and Gleitman found that reaction time was independent of the number of nontargets (i.e., flat slopes of reaction time as a function of display set size) in the between-category conditions for both target-present and target-absent trials. This was not the case for within-category conditions, which exhibited significant nonzero slopes for both present and absent trials. What is even more striking is that the results for the stimulus "O" mirrored those of the other stimuli, even though the only difference between the two category conditions was the name "O" was given by the experimenter (i.e., the character "O" was physically identical in the between- and within-category conditions).

Jonides and Gleitman interpreted this result as indicating that the difference in reaction times was not mediated by simple physical differences between the target and distractors. They further suggested that categorizing alphanumeric stimuli requires less processing capacity than identifying them. Such a hypothesis would make sense if the membership of a stimulus in an alphanumeric category is defined by fewer features than its unique identity. If this were true, then perceptual processing (e.g., feature extraction) would only need to proceed to the point needed to categorize the stimulus.

This hypothesis, termed the "partial processing hypothesis" (Jonides and Gleitman, 1976), has received some support. Jonides and Gleitman (1976) further hypothesized that the partial processing which occurs on a between-category trial results in a spatial tag allowing further analysis (identification). Thus, localization of the target in between-category conditions can be done independently of display set size (since the spatial tag allows the category-discrepant item to "pop out" of the display, eliminating the need to search the display). This account is consistent with the "oh-zero" effect described above if it is assumed (as do Jonides and Gleitman) that the character "O" contains some features of both the letter and digit categories. The subject could then choose to extract the "digit" features contained in the "O" if the background is made up of letters, and the "letter" features if the background is made up of digits. This does, however, assume that the subject's feature-extraction mechanism is flexible and amenable to modification by instruction and/or practice.

Jonides and Gleitman (1976) reasoned that subjects who were presented with between-category catch trials (containing a digit other than the digit specified as the target) would be forced to identify the digit rather than responding on the basis of category membership. Subjects were run in a between-subjects design, the three groups being assigned to the between-category, within-category, and the "modified" between-category (25% catch trials) conditions. The display consisted of two, four, or six stimuli placed around the circumference of an imaginary circle, with two possible targets specified on each trial, i.e., the memory set size was two (hereafter, "memory set size" and "M" will be used interchangeably).

Jonides and Gleitman also reasoned that subjects in the modified between-category condition should show RT functions (against display set size) with the same slope as subjects in the between-category condition, but much lower than subjects in the within-category condition. This follows from the idea that "pop-out" (the spatial tag) eliminates the need to search the display - even if the digit is the wrong one, it will still be localized independently of display set size. Further, modified between-category subjects should also show elevated RT's at all display set sizes with respect to between-category subjects, since a further memory comparison stage is necessary to reject false "targets". There is no reason that this factor should interact with display set size, since subjects should get the "spatial tag" at all display set sizes. These predictions were upheld.

Gleitman and Jonides (1976) provided further evidence for the partial processing hypothesis by demonstrating a cost associated with shallower processing. Gleitman and Jonides replicated the "oh-zero" effect in a letter background, and later looked at recognition memory for the distractors as a function of category condition. The within-category group had significantly greater recognition scores than the between-category group, indicating only "partial" processing in the between-category group. Recognition scores showed no correlation with RT in the within-category group, and no correlation between "fast" and "slow" stimulus cards. This latter result rules out the alternative explanation that the superior performance of the within-category subjects on the recognition task was an artifact of having more time to process the stimuli. Gleitman and Jonides further demonstrated that between-category subjects did not need to identify or compare the target in the display to the target in memory by showing that nine out of ten of the between-category subjects responded "present" on a final catch trial on which they were presented with a digit not in the memory set. The RT's on this trial were comparable to normal between-category trials, suggesting that subjects used the same strategy (respond on the basis of category membership) on the two kinds of trials.

There is, however, evidence that categorization follows (and indeed is based on) identification. Dick (1971) found that categorization time for a single letter was equal to the time to name the letter plus a constant. White (1977, Experiments 1-4) replicated and extended Dick's results, finding faster identification than categorization responses when target stimuli were presented alone or in various backgrounds and for various durations, ranging from 100 to 500 ms. In Experiment 1, subjects responded with the name (identification) of a cued stimulus (presented in a blank field or in between- and within-category backgrounds) or with its superordinate category (categorization). In Experiments 2 and 3, subjects "searched" (the target was localized by a bar marker) for a particular member of a category (identification) or for any member of the category (categorization). The paradigm in Experiment 4 was essentially the same as in Experiments 2 and 3, except that the target was not localized. White found categorization times which were 80-100 ms slower (across the different experimental paradigms) than identification times for stimuli presented in between-category backgrounds.

Nickerson (1973) also reports evidence that categorization of an alphanumeric character requires that the character be identified first. Nickerson (1973, Experiment 1) found accuracy to be no better than chance on "implicit categorization" of degraded stimuli that had been named incorrectly (i.e., an incorrect identification response was a member of the same category as the stimulus no more often than would be expected by chance). If subjects classify characters as a prelude to full identification, one might expect the subject to classify the stimulus "correctly" more often than chance on those trials where the identification response was incorrect. Further, when Nickerson (1973, Experiment

2) asked subjects to classify the same stimuli as digits or letters ("explicit classification"), performance was even poorer than implicit classification performance in the first experiment. If subjects were able to classify a stimulus without (or prior to) identifying it, one would expect better performance on an "explicit" classification task.

Faced with the above evidence indicating that categorization was not faster than identification, some theorists (e.g., Deutsch, 1977; Taylor, 1978; Gleitman and Jonides, 1978) renounced the partial-processing hypothesis in favor of various "semantic" (i.e., late-selection) alternatives. At the highest level, these viewpoints are essentially the same - I will consider the viewpoint put forward by Gleitman and Jonides (1978) as a paradigmatic example. Gleitman and Jonides (1978) argued that subjects need to be able to "set" themselves for the target category at a "conceptual" level. On this view, the display stimuli would be encoded to the level of category membership and compared with the target category (e.g., "letter") in memory. Gleitman and Jonides (1978, Experiment 2) asked subjects to search for a memory set containing a digit and a letter in a within-category condition for 112 trials (i.e., in appropriate within-category distractors for whichever one was the target on that trial). On the 113th trial, subjects saw an unexpected between-category trial. The slope of the unexpected between-category RT function was not significantly different from that of the within-category condition.

In Experiment 1, where the memory set consisted of two members from the same category, the slope of the unexpected RT function was significantly different from the within-category slope, and not significantly different from the slope in a "pure" between-category condition. Gleitman and Jonides suggested that abstract letter (identity) and category codes were available to the subject, and that there were "lower code overlaps" between targets and distractors in the between-category case, resulting in faster memory comparison times. Gleitman and Jonides did not explicitly state whether they believed processing in the between-category condition was done in parallel - only that category codes resulted in "...the symptom most characteristic of the category effect, lower search times per item in a between- than in a within-category condition" (p. 364). It is difficult to know whether the between-category, target-present slopes were low enough to indicate a parallel process, as Gleitman and Jonides reported the mean slope (13.0 ms/item) collapsed across target presence/absence. In any case, category codes could not reliably tell the subject whether a target was present or not in Experiment 2 (since the memory set contained representatives of both categories).

Gleitman and Jonides (1978) argued that the semantic explanation of the category effect could explain the "oh-zero" phenomenon as well. Subjects in between-category conditions could choose to encode "O" as a digit when the background was composed of letters, and as a letter when the background was composed of digits. Thus, "O" would show the same comparison time advantage as other members of its category. The semantic explanation also accounts for some of the other demonstrations of the category effect described above. First, the results of the Jonides and Gleitman (1976) modified between-category condition can be understood in this framework. Recall that Jonides and Gleitman argued that subjects could make use of a "spatial tag" to locate a between-category target. The modified between-category condition (i.e., 25% of the target present trials were catch trials) resulted in longer response times relative to the between-category condition because the presence of catch trials required subjects to do more processing to identify the target (according to Jonides and Gleitman). According to the semantic argument, subjects could locate the appropriate item based on its category code, but would then need to perform another comparison based on identity to be certain that the category-discrepant item was indeed the target (resulting in a longer response times at all display set sizes).

Further, catch trial effects like those reported by Gleitman and Jonides (1976) follow naturally from this account. If subjects are set to compare by category, any display element from the target category will initiate a "present" response. Further, subjects who are comparing stimuli in memory at the category level (between-category condition) should not be able to recognize specific targets (identity-level information) on a subsequent recognition test. In contrast, within-category subjects (who have been performing memory comparison at the identity level) should show better recognition of specific targets. Recall that this was the result obtained by Gleitman and Jonides (1976).

It would seem that this semantic explanation accounts for the category effect rather well. There is, however, a some evidence which suggests otherwise. There have been several reported failures to replicate the oh-zero effect. Duncan (1983) found positive and identical slopes of the RT vs. display set size functions for between- and within-category "O" search. White (1977, Experiments 5-7) attempted to replicate the oh-zero effect using six-element arrays. White found that "O" was detected faster in a background of digits than letters, regardless of its designation as "oh" or "zero". He argued that "O" should be detected faster in a background of digits than letters because it is more similar to letters than digits (subtending a wider visual angle than most digits). If it is the case, however, that a stimulus must be named before it can be categorized, and if it is harder to identify "O" in a letter background, subjects could still be responding based on the category that "O" has been assigned.

Further, results from Sperling, Budiansky, Spivak, and Johnson (1971), usually taken as support for a category effect, suggest the involvement of featural differences as well. Sperling et al. found that search with instructions to search for "any digit" was as accurate as search when the subject knew the particular digit which would appear (the task was to locate the digit). If "O" is indeed detected better in digits than in letters because of its visual angle (as White, 1977 argues), then it follows that it should be more difficult to detect in letters than the other (1-9) digits, since it would be more similar to letters than digits on this dimension.

In fact, performance on trials where "0" was the target was much poorer than performance with the other (1-9) digits. This was true with both "particular digit" [p(correct loc.) = .019 vs. mean of .514] and "any digit" [p(correct loc.) = .011 vs. mean of .522] instructions. The mere fact that the localization accuracy varies from .011 to .691 for the individual stimuli in the "any digit" condition also suggests that it is not category membership, but something specific to the individual stimuli (e.g., features) that mediates performance. Such an argument must be qualified, however, by the possibility that category codes must be derived from identity codes, in which case more easily derivable/more veridical identification for some stimuli might produce the observed differences.

Additionally, several studies have manipulated both category difference and target-background similarity (confusability). Corcoran and Jackson (1977, Experiment 3) pitted category difference against target-background confusability in a detection task with two, four, or six stimuli in the display. Subjects saw a single target ("C", "A", "6", or "4") on each trial in a background of straight letters, curved letters, or curved digits (only curved digits were used because there are too few straight ones). There was a main effect of target-background similarity, but no main effect of category difference. However, one must exercise caution when interpreting these results, since Corcoran and Jackson varied the target category from trial to trial. Taylor (1978, Experiment 4) has demonstrated a category effect (when the target category is the same for a block of trials) which was abolished when the target category changed from trial to trial.

There is, however, one study where physical differences were roughly controlled (in a manner analogous to that in Corcoran and Jackson's study) and a category effect was obtained. Ingling (1972) had subjects scan rows of printed characters for 3, B, 7, or Z in either a digit (2, 4, 5, 6, 8, 9) or letter (J, S, G, A, C, P) background. A single target was specified prior to each trial. According to Ingling, the digit and letter backgrounds were selected to "closely resemble" each other. Briefly, scanning time (sec/item) was significantly faster for both targets in a featurally-matched pair (e.g., 3 and B) in between-category conditions than in within-category conditions. These results must be tempered by the fact that Ingling's paradigm was quite different from the tachistoscopic paradigm typically used to investigate the category effect.

Krueger (1984) reports particularly convincing evidence that the category effect is due to physical feature differences. Krueger matched both targets and distractors as closely as possible on their constituent features (see Table 1). Krueger's subjects searched for a single target in displays of two, four, or six characters. Slopes of RT as a function of display set size were no smaller on between- than on within-category trials. Further, Krueger reports that the same type font produced a category effect in another study he conducted which did not control for featural differences.

Table 1

Targets and Distractors Used in the Present Experiments and by Krueger (1984)

	Targets	Distractors
Krueger (1984)	5 6	1 2 3 4 7 9
	S G	L Z B K J P
Experiment 1	2 3 4 6 7 8 9	C E F H J K L N P R T U V X Y
	A B D G M W Z	C E F H J K L N P R T U V X Y
Experiment 2	A C D H M R T U	1 2 3 5 6 7 9
	A C D H M R T U	L Z B S G J P
Experiment 3	A C D H M R T U	1 2 3 5 6 7 9

Critical Analysis

How are we to reconcile the large number of divergent findings reported above? When a memory set size of one is used, some studies have failed to find a category effect in visual search (Corcoran and Jackson, 1977; Krueger, 1984; White, 1977, Experiment 7) whereas other studies using a memory set size greater than one (e.g., Egeth, Jonides, and Wall, 1972, Experiments 3-5; Jonides and Gleitman, 1976; Gleitman and Jonides, 1976; Gleitman and Jonides, 1978) have found clear evidence of a category effect. Note that the studies which have failed to find a category effect and have only used a memory set size of one are also the studies which argue that the category effect is due to "uncontrolled physical feature differences". Thus, we have a confounding of memory set size and manipulation of target-background similarity.

Because of this confounding, two cases need to be considered, i.e., the case where no physical differences obtain between categories and the case where physical differences do obtain. To begin, consider the case where no physical differences obtain. When the subject is presented with an alphanumeric array, identity and category codes are produced (perhaps in parallel, as Posner (1970) suggests) with the identity code being available for response sooner. This reasoning makes sense, given the evidence that identification seems to be faster than categorization. If subjects can choose to use either the identity or category "code", I suggest that they will choose to use the category code when M is larger than one (since the added memory load should induce subjects to group the memory set as a single item, e.g., "letter"), and the identity code when M is equal to one (since it arrives first, and the subject need only compare the display with a single item, i.e., the target).

Implicit in the above view is the notion that the locus of a "conceptual" category effect (that is, a category effect obtained when M>1 which is not the result of gross featural differences) would be at memory comparison, and not encoding or "display search". I am assuming a parallel, hierarchical perceptual system whose only limitations are due to "crosstalk" and confusability within the system (see Pollatsek and Digman, 1977, for evidence supporting the notion of dependent channels in visual search). Although identity information would be produced in parallel for the whole display (as would category information), it seems that memory comparison based on category information would be more likely to be done in parallel than memory comparison based on identity information. This might be expected if memory comparison based on semantic category codes were less likely to result in "outcome conflict" (cf., Navon, 1986; Navon and Miller, 1987) by virtue of having less "code overlap" (cf., Gleitman and Jonides, 1978) than identity codes. Thus, an account which locates the category effect at memory comparison can explain why no category effect is obtained with a memory set size of one whereas a category effect is obtained with a memory set size greater than one when physical differences have been equated.

Next, assume that physical differences do obtain between the two categories. Then, if one assumes a hierarchically-organized perceptual system (e.g., Selfridge, 1959), with feature detectors at the lowest level, it makes sense that simple featural information would be available for use before identity or category information. Also, if feature(s) exist which are common to the target set and not to the distractor set, subjects could and probably would search for said feature(s) quickly and in parallel (cf., Treisman and Gelade, 1980; Treisman and Souther, 1985). Thus, search based on features might be expected to be done in parallel <u>and</u> to be faster than search based on category information.

Notice that this is not a restatement of the partial-processing hypothesis - there is no claim that the stimulus is classified based on these features, only that the response may be based on their presence/absence. Since the features which define a "present" response in between-category conditions would be the same features which uniquely define the target category, subjects would, in effect, be classifying the target as a member of the target category. This is logically different, though, from first classifying a target based on these features, and then responding based on its category membership. Recall that the partial-processing hypothesis was an attempt to explain faster between-category RT's by recourse to the notion that less processing (feature extraction) was required to categorize a stimulus than to identify it. The studies by Dick (1971) and White (1977) described above argue strongly that categorization does not take less processing than identification. This argument should be qualified, however, by the possibility that the character sets used by Dick and

White did not contain the necessary featural differences to enable "partial processing". In any case, if featural differences exist in the between-category case and not in the within-category case, then "uncontrolled physical feature differences" is a reasonable explanation of the category effect. However, if such features do not exist (because of high target-background confusability), it seems that subjects might then need to use either identity or category information.

There is some evidence supporting the notion that subjects can search for features particular to the target faster than they can search for something common to the target category (either common features or category codes). Several studies (e.g., Brand, 1971; Sperling et al., 1971) have found that search for any member (e.g., "any digit") is roughly as fast or as accurate as search for a specific member (e.g., "3"). However, Taylor (1978), Hock et al. (1985), and Schneider and Shiffrin (1977, Experiment 2, see Figure 5, p. 19) have failed to replicate this finding. Most importantly, these three studies compared performance under the two search conditions (M=1 vs. M>1) within subjects. This allows an unambiguous comparison of performance in the two conditions.

I will now consider each of these studies in turn. Hock et al. (1985, Experiment 2) found between-category search with "any member" instructions that was significantly slower than search for a single specified target. Taylor (1978, Experiment 3) estimated display set size one "intercepts" from his data, finding higher "any member" N=1 intercepts for all eight subjects (for both positive and negative responses) in the experiment. Presumably, the difference between the two search conditions when N=1 reflects the difference between identification and categorization of a single character. Unfortunately, Taylor did not test the significance of this difference.

Schneider and Shiffrin (1977, Experiment 2) ran a slightly different version of the standard between-category condition, in which the display was pre- and post-masked, and subjects were highly practiced (consistent-mapping condition). For all practical purposes, however, the paradigm is the same. Schneider and Shiffrin argued that there was no effect (on RT) of increasing the memory set size (from one to four items) in their consistent-mapping conditions, and that subjects had developed an "automatic attention response" which obviated the need to search the display or scan through the memory set. Unfortunately, they did not report the means in tabular form, nor did they report any significance tests of the difference between the two memory set size conditions. However, it appears from their Figure 5 (p. 19) that there is an increase of roughly 50 ms from M=1 to M=4 at all display set sizes. Although there are other possible explanations, I would like to suggest that the difference between the M=1 and M>1 between-category conditions in these three studies reflects the difference between identification (based on a few distinguishing features) and categorization.

To summarize, when simple featural information does not sufficiently differentiate targets and distractors and M=1, I believe that subjects may opt for a serial comparison (based on "abstract" identity codes) of the memory set to the display. This situation would be expected to obtain in those experiments described above where target-background features were manipulated and no category effect was found. Such a serial strategy would be quite costly, however, with increases in memory set size and display set size. Specifically, under between-category conditions where featural information does not differentiate the targets and distractors and where the memory set contains more than a single item, subjects might choose to "wait" for the category information, allowing comparison of the memory set as a category (a single comparison) to the category "codes" from the display in parallel.

Given this sort of reasoning, it seems desirable to ascertain under what stimulus and task conditions data indicative of the category effect are obtained. Experiment 1 was run to obtain data on two memory set sizes within subjects, in an attempt to replicate the RT differences across memory set size found in Taylor (1978), Hock et al. (1985), and Schneider and Shiffrin (1977). In Experiment 2, target-background features were controlled as in Krueger (1984), but the memory set size was varied. This resulted in the removal of the confounding of confusability/feature matching and memory set size mentioned at the beginning of this section. If subjects base their search on identity codes when M=1 and category codes when M=4, then the M=1, between-category condition should replicate Krueger's study (no category effect and serial search), but the M=4 condition should not (i.e., it should show a category effect and parallel search).

Finally, Experiment 3 was run to obtain potentially converging evidence bearing on the locus of the category effect from false alarms on catch trials. Subjects in both M=1 and M=4 between-category conditions saw catch trials containing either a "category" foil (same category as the target set, but featurally discrepant from the memory set for that trial) or a "feature" foil (different category from the target set, but featurally similar to the memory set for that trial). If subjects in the M=1 condition are paying attention to identity information, they should false alarm to the feature foil, but not to the category foil. Likewise, if subjects in the M=4 condition are paying attention to category information, they should false alarm to the category foil, but not to the feature foil.

General Description of Analyses

In experiments that have obtained differences between within- and between-category search times, there is also a different pattern of response times as a function of display set size that suggests a qualitative difference in the process. More specifically, within-category conditions typically exhibit positive slopes conforming to a roughly 2:1 ratio (absent:present), while between-category slopes for present responses are typically not different f om zero. Between-category absent responses should show an increasing trend, assuming variable examination times for individual stimulus items (see Egeth et al., 1972, p. 679).

The zero slope for positive responses is commonly taken to indicate parallel processing based on something about the target category (semantic or featural). Under a parallel model, absent response times could exhibit either a flat (zero slope) function, or a negatively-accelerated increasing function. If subjects set a "deadline" after which they respond "absent" if they have not detected a target, the function for absent responses will be flat. If subjects examine all items in the display before responding "absent", then the function should increase as display set size increases. In the former case (both "present" and "absent" slopes equal to zero), the ratio of slopes would be undefined. In the latter case ("present" slope essentially zero, "absent" function increasing) the ratio of slopes would be greater than 2:1. In either case, using linear parameters to describe a parallel process is not advisable. We can, however, use them to rule out a serial model. Nonzero positive slopes in a 2:1 ratio have been taken to indicate some sort of serial, self-terminating "search" based on something other than category information (but see Townsend, 1971, for a discussion of the difficulties inherent in distinguishing serial from parallel processes). I will take this as my starting point also.

CHAPTER 2

EXPERIMENT 1

The reasons for running Experiment 1 were to obtain data on the category effect using two memory set sizes within subjects, and to attempt to obtain the RT differences (across memory set size) found in Taylor (1978) and Hock et al. (1985). Experiment 1 also affords the opportunity to demonstrate a category effect with the type font and apparatus used in these experiments.

Method

Subjects

Eight undergraduates at the University of Massachusetts participated for credit.

Stimuli and apparatus

The stimulus set used appears in Table 1. The between-category targets were digits, the within-category targets were letters, and the distractors were letters. The stimuli were software-generated and were presented on a Hewlett-Packard X-Y point-plotting display (Model 1332A) with a P31 phosphor (decay to 1% intensity .25 ms after display offset), driven by a Zenith Z100 microcomputer. The characters were .63 cm in height and .40 cm wide, and were viewed from a distance of approximately 76 cm. Subjects were run alone in a dimly lit room.

Two, four, or six stimuli were presented in a subset of the 12 clock positions around the circumference of an imaginary circle subtending approximately 3.4° of visual angle. The distractors were selected randomly (without replacement) and placed randomly with the following constraints. If the display set size was two, the distractors appeared diametrically opposite each other. If the display set size was four or six, the distractors were evenly spaced. For example, stimuli might appear at 2, 4, 6, 8, 10, and 12 o'clock for display set size six, and 3, 6, 9, and 12 o'clock for display set size four. This method kept the visual angle constant across changes in display set size, and was an attempt to minimize effects of increasing density. On half the trials in a block, a target replaced one of the distractors.

Design

The design was a 2 X 2 X 2 X 3 within-subjects design. The within-subjects factors were memory set size (one vs. four), category condition (between vs. within), target presence/absence, and display set size (2, 4, 6). Half the subjects were presented with the memory set size one trials first, the other half with the memory set size four trials first. Two of the four subjects saw the between-category trials first and the other two saw the within-category trials first. Thus, category condition was nested in memory set size. Within each block of trials, all subjects saw an equal number of trials of the three levels of display set size (2, 4, and 6).

Procedure

Subjects were run in a single one-hour session of 576 trials in four blocks of 144 trials (with a short break after each block), following a practice block of 36 trials. Subjects saw all the trials associated with one memory set size in the first two blocks, and all the trials associated with the other memory set size in the second two blocks. Prior to each trial, subjects were presented with the target(s) they were to search for. Different targets were selected for each trial (as opposed to using the same memory set for a block of trials) in order to encourage the use of a "category" comparison strategy, if indeed it is possible to search based on category membership. That is, blocking the memory set might encourage search based on featural or identity information, thus obscuring any ability the subject may have to search based on category information. A predisposition to search based on featural or identity information might result in something more like serial search in the M=4 condition, which would in turn result in the (possibly false) conclusion that subjects cannot search based on category information.

Subjects pressed a response button (different from those designated for "present" and "absent" responses) to initiate the trial. After a 500 ms delay, the display was presented for 175 ms, after which subjects made a speeded response indicating the presence or absence of a target. Half the subjects responded "present" with their dominant hand while the other half responded "absent" with their dominant hand. They then received feedback on the accuracy of their response. Subjects were asked to "be as accurate as possible, and within that constraint, respond as quickly as possible".

Results

Reaction times

Mean RT for correct responses and error proportions were calculated for each subject. Reaction times deviating more than ± 2.5 standard deviations from the mean for their particular (memory set size X category X display set size X target presence/absence) cell were discarded. As such, the reported means are from "trimmed" data. Functions of RT plotted against display set size for between- and within-category search appear in Figures 1 and 2, respectively. The data were examined by analysis of variance, with memory set size (one vs. four), category condition (between vs. within), target presence/absence, and display set size (2, 4, 6) as within-subjects factors. The main effects of category and memory set size reached significance [$F(1,7) = 75.62$, $p < .001$ and $F(1,7) = 21.19$, $p < .01$, respectively] as did their interaction [$F(1,7) = 77.65$, $p < .001$]. See Appendix B for the full Anova table. The between-category, target-present RT's were further tested by analysis of variance (memory set size X display set size) for the expected memory set size difference (based on the results from Schneider and Shiffrin, 1977, Taylor, 1978, and Hock et al., 1985). The difference was in the right direction (i.e., M=4 > M=1), but did not reach significance [$F(1,7) = 2.80$, $p > .10$].

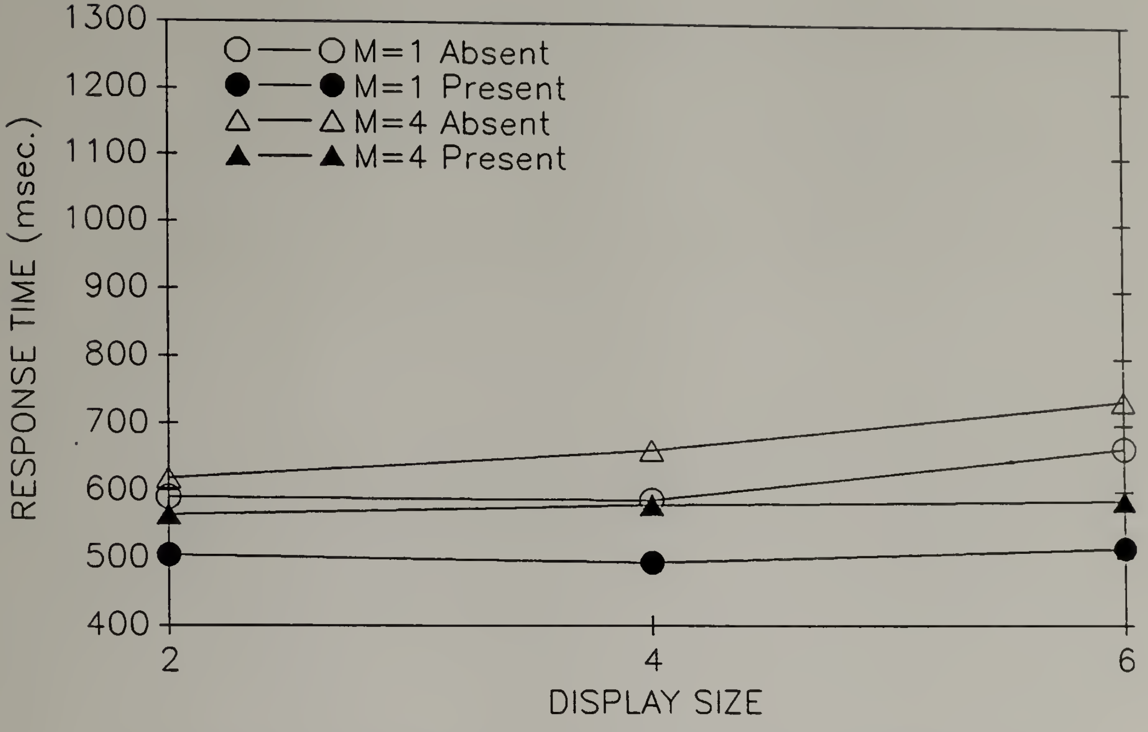

Figure 1. Response time as a function of display size for the between-category conditions in Experiment 1.

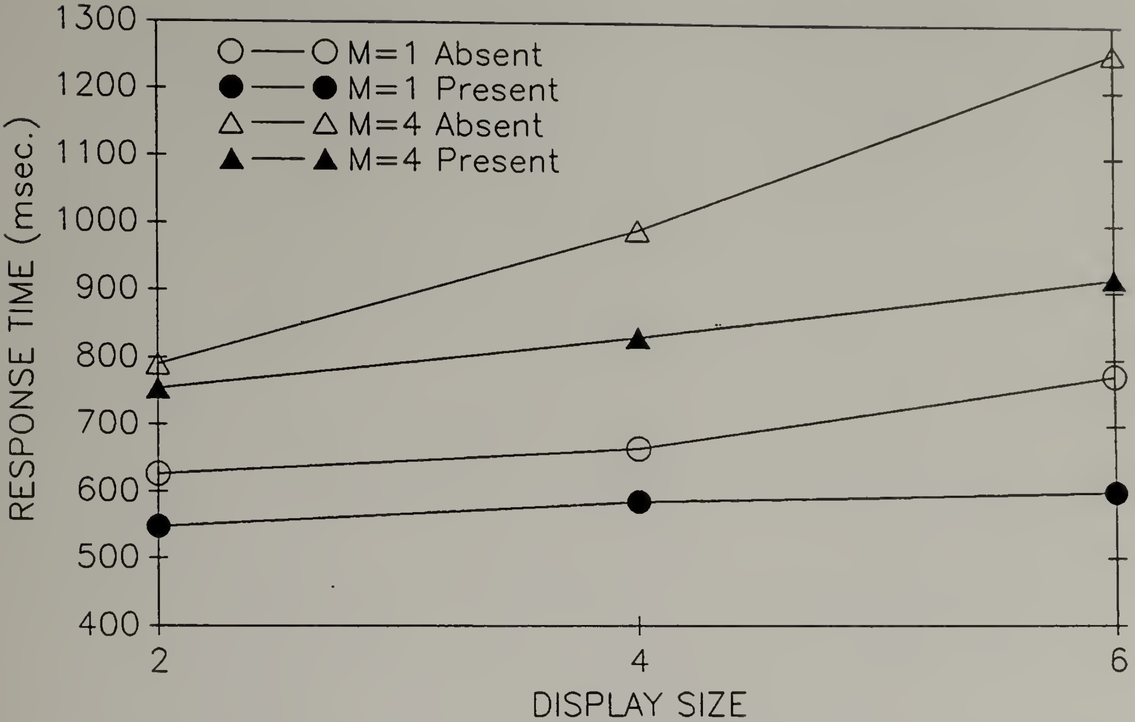

Figure 2. Response time as a function of display size for the within-category conditions in Experiment 1.

In order to further explore the effect of increases in display set size on search behavior in the various conditions, straight lines were fit to functions of RT against display set size for each subject. Mean slopes and intercepts for the best-fitting regression lines appear in Table 2. The individual subject slopes were subjected to analysis of variance, with memory set size (one vs. four), category condition (between vs. within), and target presence/absence as within-subjects factors (see Appendix B for the full Anova table). Again, the main effects of category and memory set size reached significance [$F(1,7) = 16.15$, $p < .01$ and $F(1,7) = 17.20$, $p < .01$, respectively] as did their interaction [$F(1,7) = 7.83$, $p < .05$].

Post-hoc dependent t-tests (adjusting the alpha level using the Bonferroni procedure) were performed to further investigate the category condition X memory set size interaction. Since the three-way interaction of these two factors with target presence/absence was not significant, present and absent slopes were averaged for the analysis. The mean slope for a memory set size of four was significantly greater than that for a memory set size of one in the within-category condition [$t(7) = 3.53$, $p < .01$], but not in the between-category condition [$t(7) = 1.68$, $p > .05$]. The mean slope for within-category trials was significantly greater than that for between-category trials in both memory set size one [$t(7) = 3.72$, $p < .01$] and memory set size four [$t(7) = 3.53$, $p < .01$] conditions.

Table 2

Mean Slopes (Milliseconds/Item) and Intercepts (Milliseconds) of Best-Fitting Regression Lines for Functions of RT Plotted Against Display Set Size in Experiment 1

		Within-Category Search		Between-Category Search	
		Target		Target	
		Present	Absent	Present	Absent
Slope	M = 1	14 ± 17	37 ± 15	3 ± 7	19 ± 7
	M = 4	41 ± 39	117 ± 49	6 ± 5	29 ± 14
Intercept	M = 1	523 ± 72	540 ± 90	493 ± 67	538 ± 77
	M = 4	670 ± 106	547 ± 113	554 ± 96	555 ± 110

Note. Slopes and intercepts appear with 95% confidence intervals.

Errors

Mean error proportions are shown in Table 3. Overall, error rates were somewhat higher than those obtained in previous studies. As might be expected, errors tended to increase with increasing display set size, especially in the within-category, target-absent conditions.

Table 3

Error Proportions for Experiment 1

		Within-Category Search			Between-Category Search		
		display set size			display set size		
	Target	2	4	6	2	4	6
M = 1	Present	.081	.124	.127	.040	.066	.091
	Absent	.050	.066	.170	.056	.045	.077
M = 4	Present	.219	.296	.246	.066	.066	.050
	Absent	.035	.116	.332	.030	.010	.109

Discussion

There are four aspects of the data which are noteworthy. First, there was an effect of category in both M=1 and M=4 conditions in the present experiment. Between-category slopes of reaction time as a function of display set size were comparable in magnitude to those found by Egeth et al. (1972, Experiment 3) and significantly lower than those in the within-category conditions. Further, the increase in error rates with increases in display set size was more pronounced in the within-category conditions, particularly in the M=4, target absent condition (see Table 3). Although there is no obvious speed-accuracy trade-off in these conditions, it could be argued that the slopes in these conditions would have been even larger had subjects been able to maintain accurate performance.

Second, the ratio of absent to present slopes was approximately 2:1 in the within-category, M=1 condition, and was far greater in both between-category conditions. It is difficult to interpret the ratio of slopes in the within-category, M=4 condition, due to the high and variable (across display set size) error rates in this condition. However, based on the M=1, within-category condition, it appears subjects were searching in a serial self-terminating fashion in the within-category conditions, and in a different manner in the between-category conditions.

Since the target-present slopes for the between-category conditions are quite small (see Table 2), the likely interpretation is that subjects were searching in parallel in these conditions. The between-category target-absent slopes were somewhat higher. Egeth et al. (1972) also found relatively large (28.1 ms/item) absent slopes early in practice (Day 1 of Experiment 4) which decreased to 4.1 ms/item by Day 4 of Experiment 4. Egeth et al. suggested that, early in practice, subjects would recheck the display in series to verify that no target was present if a "present" decision had not been made after some interval. With practice, their argument goes, the subject learns to "trust" the initial parallel examination of the stimuli. A similar process could be operating in the present experiment. There is, however, no reason to assume that this is the case, since variation in stimulus examination times would result in an "absent" function which increases with N.

Third, the finding of similar slopes (in a within-subjects design) by Taylor (1978) and Hock et al. (1985) for one vs. several targets in between-category search was upheld in the present experiment. Additionally, the mean RT in the between-category, M=1 condition was lower (albeit not significantly) than the mean RT in the between-category, M=4 condition (see Figure 1), which is consistent with the results of Taylor (1978) and Hock et al. (1985). It is tempting to conclude from this stable (across different experiments) RT difference that different information is being utilized in the two between-category conditions. More specifically, as argued above (and in Appendix A), use of featural information should result in lower RT's at all levels of display set size relative to those resulting from the use of category information. However, since featural differences across category conditions were not controlled for and the difference is not significant, it is possible that featural information is being used when M=4 as well as when M=1. This possibility will be tested further in Experiments 2 and 3.

Finally, the error rates in the present experiment were rather high, particularly in the within-category conditions (see Table 3).

It could be argued that such high error rates (in the within-category conditions) make interpretation of the corresponding reaction times difficult or impossible. Still, the massive difference in error rates between the two category conditions when M=4 suggests that something is available (be it featural or category information) in the between-category conditions which results in more efficient search behavior than that found in the within-category conditions.

CHAPTER 3

EXPERIMENT 2

Although the results of Experiment 1 indicate a "category effect", it is still necessary to refine our knowledge about the locus of the effect ("early" or "late", i.e., featural or semantic). To this end, an experiment similar to that of Krueger (1984) was undertaken, with one important modification - memory set sizes of one and four were used. Presumably with featural information controlled across the two category conditions (unlike in Experiment 1), between-category targets would not be more discriminable than within-category targets, so that subjects in the between-category conditions should not be able to search based on simple featural information. Thus, it was hypothesized that subjects searching on the basis of identity information (memory set size one) would proceed serially when comparing the display to the target in memory. Memory set size four trials, on the other hand, should encourage comparison on the basis of category codes, a comparison which I have argued could be made in parallel.

Also, the procedure was modified an attempt to reduce error rates and RT variability relative to Experiment 1. First, if subjects' accuracy dropped below a preset criterion, they were encouraged to be more accurate. Second, subjects were given more practice in the task. They were run in two sessions, with the first session serving as practice.

Method

Subjects

Sixteen undergraduate and graduate students at the University of Massachusetts participated for payment.

Stimuli and apparatus

The apparatus was the same as in Experiment 1. The items were located in the 12 clock positions as in Experiment 1. Stimuli were similar to those used by Krueger (1984; see Table 1). Due to the small number of digits relative to letters, subjects searched for the same target set of letters (A, C, D, H, M, R, T, and U) in both the between- and the within-category conditions. Krueger's digit (between-category condition) and letter (within-category condition) stimuli served as distractors. The target letters were chosen so as not to allow them to be easily discriminable in any gross way (e.g., curvature) from Krueger's stimuli.

Again, it is desirable to encourage the use of a semantic strategy if indeed subjects are able to use it. If the targets could be too easily discriminated from the background (e.g., if the targets were all straight letters) subjects might be able to search on simple feature differences, and not have to resort to a category strategy. Although any target-background featural differences would be identical in the between- and within-category conditions (to the extent that Krueger's stimuli are matched for their constituent features), finding parallel search functions in both category conditions would only show that subjects can search in parallel for a few distinguishing feature(s). Since the objective of this experiment was to ascertain whether category membership can be used to search in parallel, such a result would tell us little.

Design

The design was a 2 X 2 X 2 X 3 mixed design. The within-subjects factors were memory set size (one vs. four), target presence/absence, and display set size (2, 4, 6). The between-subjects factor was category condition (between vs. within). All subjects saw an equal number of trials of three levels of display set size (two, four, and six).

Procedure

Subjects were run in two one-hour sessions of 584 trials, preceded by 48 practice trials in the first session. In addition, the first two trials in each block of 146 trials were counted as practice and were not included in subsequent data analysis (leaving four blocks of 144 trials each per session). Subjects were run in two sessions in an effort to increase accuracy levels and minimize error variance in the reaction times. Prior to each trial, subjects were shown the target(s) they were to search for on that particular trial. Subjects pressed a response button to initiate the trial. The display was presented for 175 ms, after which subjects made a speeded response indicating the presence or absence of a target. The usual speed-accuracy instructions were given. Feedback on the correctness of the response was provided after each trial. Additionally, if accuracy on any one of the six kinds of trials (target presence/absence X display set sizes two, four, and six) in a block dropped below 95%, a message was displayed, encouraging the subject to be more accurate.

Finally, between-category subjects saw a block of 60 trials at the end of the second session. This last block of trials contained ten randomly interspersed "catch" trials (containing various "foils" to which the correct response was "absent"). These trials were pooled with the data from Experiment 3 and will be discussed in detail as part of Experiment 3.

Results

Reaction times

The reaction time and error data were summarized in the same manner as in Experiment 1. Functions of RT against display set size for between- and within-category search appear in Figures 3 and 4, respectively. The data were examined by analysis of variance, with category condition (between vs. within) as a between-subjects factor and memory set size (one vs. four), target presence/absence, and display set size (2, 4, 6) as within-subjects factors (see Appendix B for the full Anova table). The main effects of category and memory set size were again significant [$F(1,14) = 6.06$, $p < .05$ and $F(1,14) = 87.17$, $p < .001$, respectively] as was their interaction [$F(1,14) = 16.74$, $p < .01$]. The RT advantage of memory set size one over memory set size four in the between-category condition (which did not reach significance in Experiment 1) was significant [$F(1,7) = 32.27$, $p < .002$]. There was no interaction of memory set size and display set size in the between-category, target-present RT's [$F(2,14) < 1$].

As in Experiment 1, straight lines were fit to functions of RT against display set size for each subject. Mean slopes and intercepts for the best-fitting regression lines appear in Table 4. The individual subject slopes were subjected to analysis of variance, with category condition (between vs. within) as a between-subjects factor, and memory set size (one vs. four) and target presence/absence as within-subjects factors. All main effects and interactions were significant at the .05 level (see Appendix B for the full Anova table).

Since the three-way interaction of category condition, memory set size, and target presence/absence was significant, the interaction of category and memory set size was first assessed by analysis of variance for present and absent slopes separately. This simple interaction was significant for both present [$F(1,14) = 5.81$, $p < .05$] and absent [$F(1,14) = 11.72$, $p < .01$] responses.

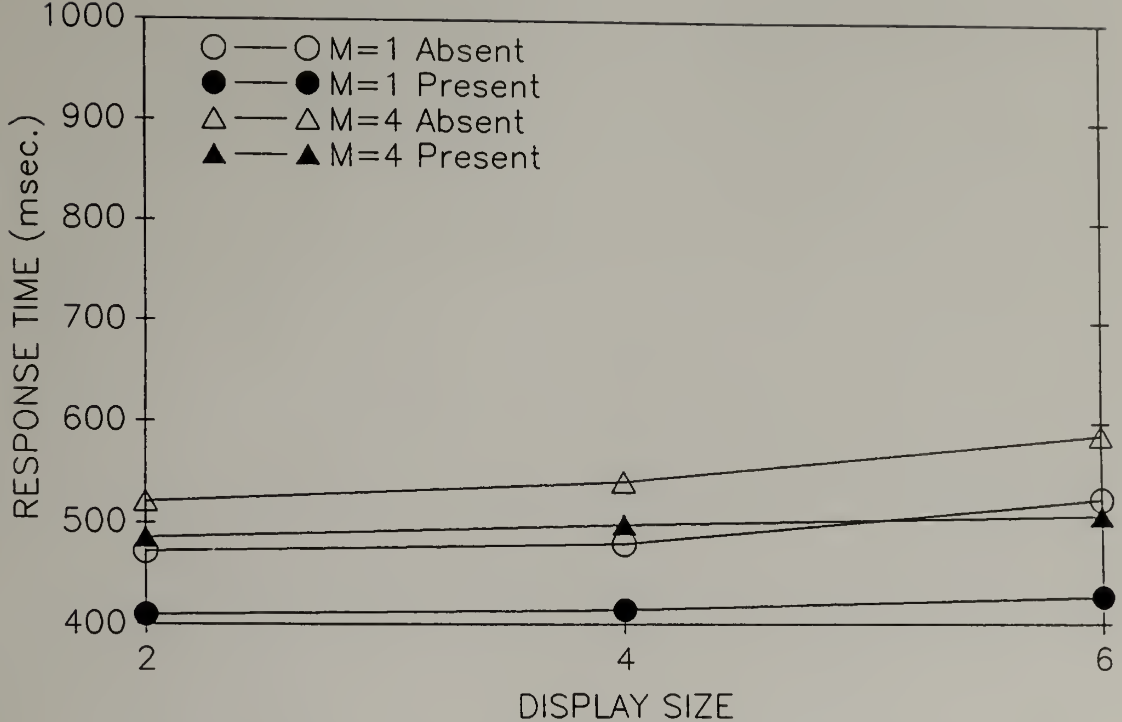

Figure 3. Response time as a function of display size for the between-category conditions in Experiment 2.

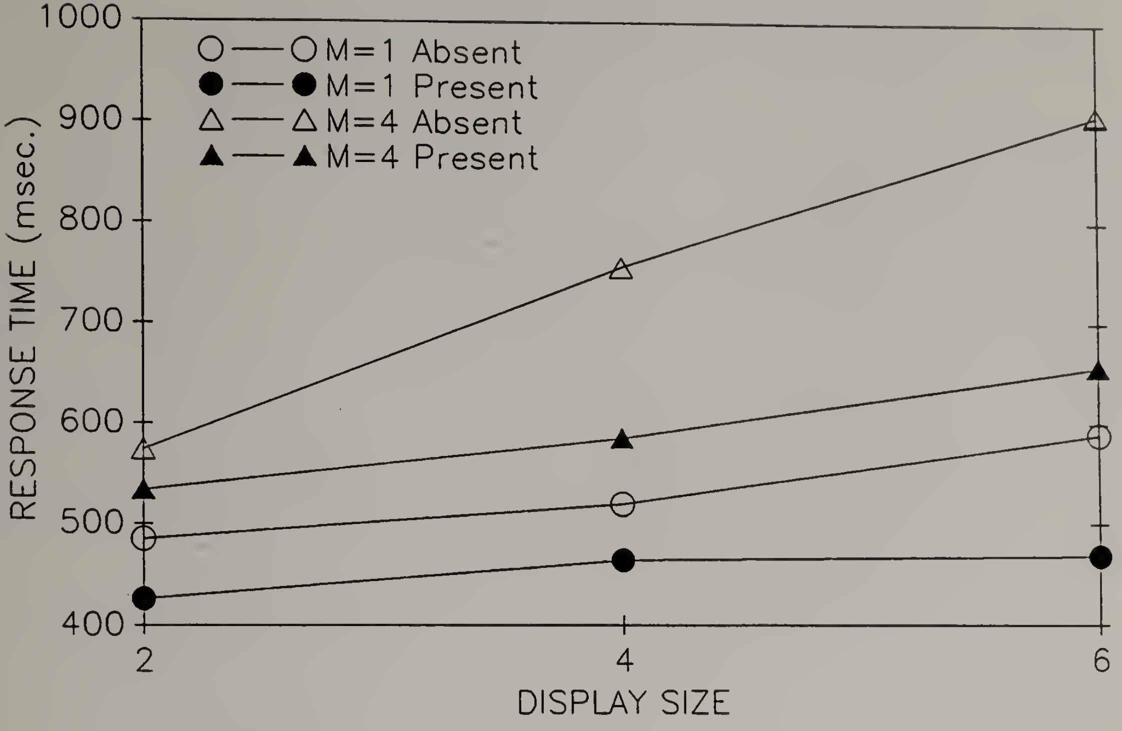

Figure 4. Response time as a function of display size for the within-category conditions in Experiment 2.

Table 4

Mean Slopes (Milliseconds/Item) and Intercepts (Milliseconds) of Best-Fitting Regression Lines for Functions of RT Plotted Against Display Set Size in Experiment 2

		Within-Category Search		Between-Category Search	
		Target		Target	
		Present	Absent	Present	Absent
Slope	M = 1	11 ± 4	26 ± 9	4 ± 3	13 ± 6
	M = 4	31 ± 21	83 ± 38	5 ± 5	17 ± 6
Intercept	M = 1	411 ± 49	427 ± 68	399 ± 33	441 ± 36
	M = 4	469 ± 61	414 ± 80	477 ± 66	483 ± 59

Note. Slopes and intercepts appear with 95% confidence intervals.

Post-hoc t-tests (again using the Bonferroni procedure, with alpha = .05/4 = .0125) were performed to further investigate the category condition X memory set size interaction. All tests were one-tailed. For absent responses, the mean slope for a memory set size of four was significantly greater than that for a memory set size of one in the within-category condition [$t(7) = 3.70$, $p < .005$], but not in the between-category condition [$t(7) = 1.86$, $p > .05$]. For present responses, the trend was the same, with no difference in the between-category condition [$t(7) = 0.37$, $p > .05$] and a marginal difference in the within-category condition [$t(7) = 2.66$, $p < .025$]. The mean slope for within-category trials was significantly greater than that for between-category trials in both memory set size one and memory set size four conditions for absent [$t(14) = 2.93$, $p < .01$ and $t(14) = 4.08$, $p < .005$, respectively] and present [$t(14) = 2.89$, $p < .01$ and $t(14) = 2.75$, $p < .01$, respectively] responses.

Errors

Mean error proportions are shown in Table 5. Overall, error rates were lower than those obtained in Experiment 1. With the exception of the M=4, within-category condition, errors did not increase substantially with increasing display set size.

Table 5

Error Proportions for Experiment 2

		Within-Category Search			Between-Category Search		
		display set size			display set size		
	Target	2	4	6	2	4	6
M = 1	Present	.035	.060	.059	.039	.022	.022
	Absent	.030	.040	.066	.012	.012	.015
M = 4	Present	.104	.154	.205	.025	.020	.027
	Absent	.015	.027	.210	.012	.020	.054

Discussion

As in Experiment 1, there was an effect of category in both M=1 and M=4 conditions in the present experiment. Between-category slopes of reaction time as a function of display set size were significantly lower than those in the within-category conditions. Also, the ratio of absent to present slopes was approximately 2:1 in the within-category, M=1 condition, and was approximately 3:1 in both between-category conditions. Again, the target-present slopes for the between-category conditions were quite small (and indeed, not different from zero in the M=4 condition, see Table 4). As before, it is difficult to interpret the ratio of slopes in the within-category, M=4 conditions, due to the rapid increase in error rates for the larger display set sizes in this condition.

The failure to replicate the Krueger (1984) finding of serial search in the between-category, M=1 condition is somewhat disturbing. Recall that featural differences across category conditions were controlled by matching the distractors on constituent features and using the same targets in both category conditions. It is tempting to conclude then, that semantic category information was all that was available, and was being utilized in both (M=1 and M=4) between-category conditions. However, as in Experiment 1, the mean RT in the between-category, M=1 condition was (significantly) lower than that in the between-category, M=4 condition, again confirming the results of Taylor (1978) and Hock et al. (1985). As argued above, featural information should arrive before category information, resulting in faster RT's at all levels of display set size. It is possible that, in between-category search, featural information is being used when M=1, and category information is used when M=4.

Thus, there are still three possible accounts of the between-category data: (1) subjects are using category information in both memory set size conditions, (2) subjects are using featural information when M=1 and category information when M=4, and (3) subjects are using featural information for both between-category, memory set size conditions. The first possibility seems to be the most likely, since featural differences should have been controlled. Still, it is possible that the target set was somehow a bit more discriminable from the digits on some dimension (e.g., visual angle), allowing search based on some simple featural difference when M=1, but not when M=4. If these featural differences exist, it is, of course, also possible (though not likely, given the significant RT difference between the two memory set conditions) that subjects in both between-category conditions are using featural information. Experiment 3 was run in order to arbitrate among these three alternative accounts of the between-category data.

CHAPTER 4

EXPERIMENT 3

Given the results of Experiment 2, it would be desirable to have some convergent evidence bearing on the level at which stimuli are selected for (early/featural vs. late/categorical) in the present paradigm. Accordingly, the purpose of Experiment 3 was to ascertain which attributes of the stimulus subjects were using in the two between-category memory set conditions. Subjects were run exactly as in the between-category conditions of Experiment 2 (except that memory set size was a between-subjects variable). Following these "normal" trials, subjects saw a block of trials with "catch" trials interspersed. The catch trials contained either a "category foil" (a distractor of the same category as the memory set for that trial, but featurally dissimilar to all members of the memory set) or a "feature foil" (a distractor of the opposite category to that of the memory set for that trial, but featurally similar to one of its members).

If subjects are attending to category information in the M=4 condition, they should show a high proportion of false alarms to category foils, but not to feature foils. Likewise, if subjects are attending to featural information in the M=1 condition, they should show a high proportion of false alarms to feature foils, but not to category foils. Thus, subjects in the two memory set conditions should show different patterns of false alarms to the two kinds of foils.

Method

Subjects

Forty-eight undergraduates at the University of Massachusetts participated for course credit.

Stimuli and apparatus

The apparatus and method of locating the items in the display were identical to that in Experiments 1 and 2. The stimuli were the same as those used in the between-category conditions of Experiment 2 (see Table 1) except for the inclusion of "B" and "4" as "foils" on catch trials. Novel foils were used on catch trials (rather than foils from the stimulus set for normal trials) to preclude alternative explanations based on repeated exposures of the foils as targets or distractors on previous trials (cf., Shiffrin, Dumais, and Schneider, 1981). In addition, the character font was altered so that the feature foil "4" was composed of the same pixels as the letter "H" minus the lower-left line segment.

Design

The design was a 2 X 2 X 2 X 3 mixed design. The between-subjects factors are memory set size (one vs. four) and type of catch trial foil (category vs. feature). The within-subjects factors were target presence/absence and display set size. All subjects saw an equal number of trials at three levels of display set size (two, four, and six) as in Experiments 1 and 2.

Procedure

Following 48 practice trials, subjects were run in a single 45 min. session of 292 "normal" between-category trials in two blocks of 146 each, plus a third block of 60 trials, 10 of which were catch trials. The first two trials of each block were discarded as practice, as in Experiment 2. The first catch trial (the primary trial of interest) occurred on the 11th trial of the third block. The memory set for the catch trials was either an "H" (M=1) or a random permutation of the letters "A", "H", "M", and "T" (M=4). Both initial catch trials (category "B" and feature "4" trials) were display set size six trials, with the same distractors (2, 3, 6, 7, and 9) appearing in the same five "non-foil" positions. Thus, the only difference between the category and feature catch trials was the identity of the foil ("B" or "4") used in that particular condition. The other nine catch trials (occurring on the same randomly interspersed trials for each subject and using the same foil as the initial catch trial) consisted of three trials for each of the three display set sizes.

Prior to each trial, subjects were presented with the target(s) they were to search for. Subjects pressed a response button to initiate the trial. The display was presented for 175 ms, after which subjects made a speeded response indicating the presence or absence of a target. Speed-accuracy instructions and feedback on the correctness of the response were provided in the same manner as in Experiment 2.

Results

False alarms

The false alarm data (pooled from Experiments 2 and 3) for the initial catch trial appear in Table 6[1]. With a few exceptions, subjects generally became aware (after receiving feedback on the first catch trial) of the catch trials after the first or second catch trial. Since the data from these trials are likely to be somewhat "contaminated", they will not be considered further. The pattern of false alarms on the initial catch trial departs significantly from what would be expected if memory set size and foil type were independent [$\chi^2(1) = 5.00$, $p < .05$].

Since the expected frequency in three of the four cells in Table 6 is less than 10, Fisher's exact probability test was also performed on the false alarm data. This procedure confirmed that the pattern of false alarms was unlikely to result by chance if memory set size and foil type were independent [$p = .009$, two-tailed]. Looking at Table 6, it is apparent that the cells most responsible for the departure from the expected frequencies are those corresponding to the two category foil conditions. While 9 of 14 subjects in the M=4, category foil condition responded "present" to the catch trial, none of the 14 subjects in the M=1, category foil condition made the incorrect response.

Table 6

False Alarm Proportions Pooled Across Experiments 2 and 3

	Memory set size	
Type of foil	1	4
Category	0	.64
Feature	.57	.50

Reaction times

The reaction time and error data were summarized in the same manner as in Experiments 1 and 2. In addition, subjects having an error rate greater than 12% in any cell or a drop in accuracy greater than 9% across display set size in any condition (memory set size X target presence/absence) were discarded. This resulted in the replacement of 6 subjects[2]. Functions of RT against display set size for both memory set sizes appear in Figure 5. The data were examined by analysis of variance, with memory set size (one vs. four) as a between-subjects factor, and target presence/absence and display set size (2, 4, 6) as within-subjects factors (see Appendix B for the full Anova table). There was no effect of memory set size, $F(1,46) < 1$. This is a departure from the results of the previous two experiments - however, memory set size is a between-subjects factor in the present experiment, making interpretation of a nonsignificant effect problematic.

As in the previous experiments, straight lines were fit to functions of RT against display set size for each subject. Mean slopes and intercepts for the best-fitting regression lines appear in Table 7. The individual subject slopes were subjected to analysis of variance, with memory set size (one vs. four) as a between-subjects factor and target presence/absence as a within-subjects factor. The main effects and their interaction were significant at the .05 level (see Appendix B for the full Anova table).

Post-hoc t-tests were performed to further investigate the memory set size X target presence/absence interaction. For absent responses, the mean slope for a memory set size of four was significantly greater than that for a memory set size of one [$t(46) = 3.08$, $p < .01$, two-tailed], but not for present responses [$t(46) = 0.75$, $p > .05$].

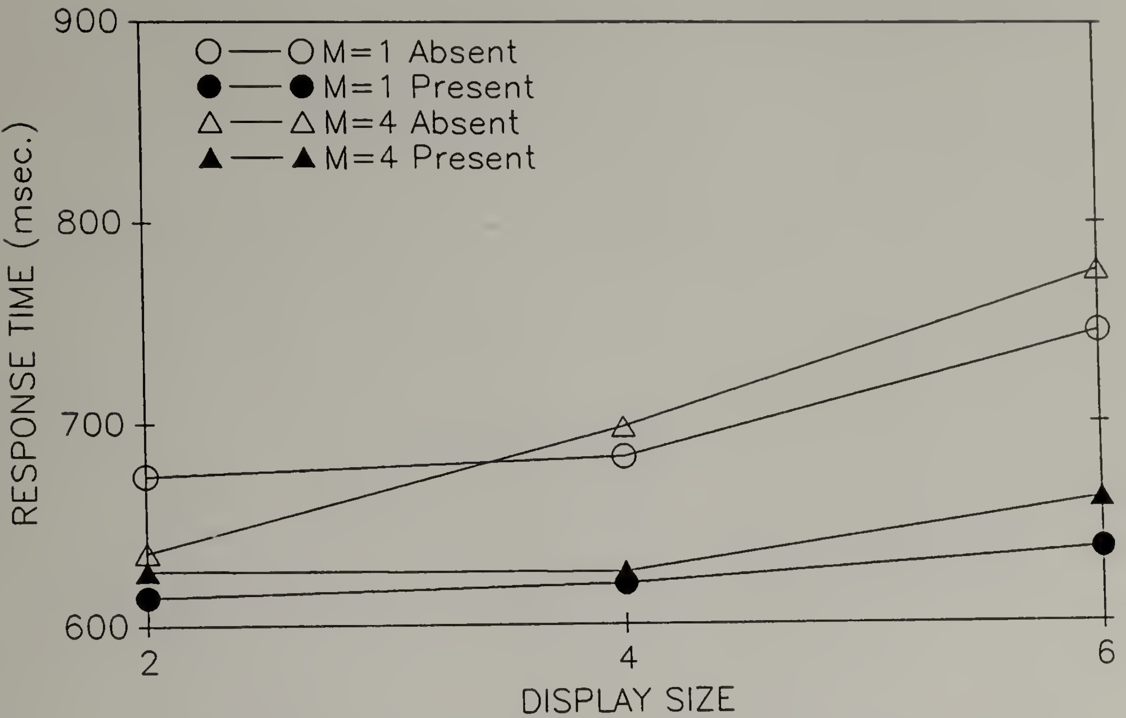

Figure 5. Response time as a function of display size for both memory set size conditions in Experiment 3.

Table 7

Mean Slopes (Milliseconds/Item) and Intercepts (Milliseconds) of Best-Fitting Regression Lines for Functions of RT Plotted Against Display Set Size in Experiment 3

	M=1		M=4	
	Target		Target	
	Present	Absent	Present	Absent
Slope	6 ± 6	18 ± 8	9 ± 5	35 ± 8
Intercept	600 ± 55	629 ± 65	603 ± 50	563 ± 35

Note. Slopes and intercepts appear with 95% confidence intervals.

Errors

Mean error proportions are shown in Table 8. Error rates were a bit higher but comparable to those obtained in the between-category condition of Experiment 2. As in the between-category conditions of Experiment 2, errors did not increase with increasing display set size.

Table 8

Error Proportions for Experiment 3

	M=1			M=4		
	display set size			display set size		
Target	2	4	6	2	4	6
Present	.049	.041	.045	.042	.023	.039
Absent	.039	.018	.033	.027	.038	.059

Discussion

It is clear from the pattern of false alarms to the category foil that subjects were utilizing different information in the two memory set size conditions. If subjects were using featural information when M=1, one would expect few false alarms to a category foil ("B") which is featurally dissimilar to the target ("H") for that trial. Indeed, none of the subjects in this condition incorrectly responded "present" to the category foil. On the other hand, nine subjects in the M=4, category foil condition incorrectly responded "present" to this same foil. This suggests that subjects were coding the memory set by category (i.e., as "letter") and were insensitive to the features of the targets ("A", "T", "H", and "M") for that trial. These results, combined with the elevated RT's found (when M=4) in Experiments 1 and 2, argue that featural information is used when M=1 and category information is used when M=4.

If all fourteen subjects in the M=4, category foil condition were searching based on category information, then all fourteen subjects should have responded incorrectly. Given the rather large slope for present responses in this condition (9 ms/character vs. 6 ms/character and 5 ms/character in the M=4, between-category conditions of Experiments 1 and 2, respectively), it seemed possible that subjects could be partitioned into two groups based on whether or not they responded incorrectly. Indeed, the five subjects who did not false alarm to the category foil had a significantly higher mean present slope (23 ms/character) than those who did (6 ms/character), [$t(12) = 1.91$, $p < .05$, one-tailed]. This result suggests that these five subjects were being more careful, either rechecking the display, and/or paying more attention to item-specific characteristics of the memory set. Thus, it appears that these five subjects were not searching based on category information.

The interpretation of the pattern of false alarms in the feature foil conditions is less clear. If subjects in the M=4 condition were responding based on category information, the digit "4" (shaped like "H") shouldn't have resulted in an incorrect "present" response. It is possible that the subjects making the incorrect response in the M=4, feature foil condition "miscategorized" the "4" as a letter, since they had been categorizing a similar form ("H") as a letter previous to the catch trial. Recall that subjects never saw the "4" previous to the catch trial, and hence had not previously categorized this form as a digit. Thus, if a subject was partitioning the stimuli in the display into two categories, this mechanism may have been "tuned" to categorize anything "H-like" as a letter.

A similar "tuning" of a search based on item-specific featural information may also account for why only eight of the fourteen subjects incorrectly responded "present" in the M=1, feature foil condition. It may be that subjects became so good at discriminating the "H" by its features that even the feature foil was not sufficiently similar to the "H" to cause more of the subjects to false alarm to it. Some of the subjects in this condition even asked whether there was "something wrong with the 'H'" or whether "a piece of the 'H' was missing".

To summarize, the data from the two category foil conditions suggest that featural information is used when M=1, while category information is used when M=4. The pattern of false alarms to the feature foil is not as conclusive, but is still suggestive. If categorization is based on identification, and the "4" is incorrectly identified as an "H", then it will be incorrectly categorized as a letter. On this argument, roughly the same proportion of subjects might be expected to incorrectly categorize the feature foil when M=4 as incorrectly identify the feature foil when M=1.

CHAPTER 5

GENERAL DISCUSSION

The most striking finding in these experiments is that subjects can search in parallel for letters in digits (but not for letters in letters) when the same letter target set is used in letter and digit backgrounds which are matched on constituent features. This finding is in sharp contrast to other studies mentioned above (most notably, Krueger, 1984) which have equated target-background confusability in between- and within-category conditions and found no difference in search rates. One obvious possibility for the departure from Krueger's findings is that the target set used in Experiments 2 and 3 was somehow more discriminable from the background than Krueger's target set. However, the target set for these experiments was chosen so as to have the same number of "straight" and "curved" members. Even with this rough "control" on target-distractor discriminability, it is still possible that the letter targets were somehow more discriminable from the digit background than from the letter background.

I believe, however, that there is still a more likely explanation for the discrepancy from Krueger's findings. Krueger (1984) manipulated not only target and background class (or "category"), but also target and background orientation (i.e., normal vs. mirror-reversed characters). In his first experiment, Krueger mixed background class and orientation <u>within</u> blocks of 24 trials. In Experiment 2, he blocked target class, target orientation, background class, and background orientation, yielding 16 (2 X 2 X 2 X 2) blocks of 24 trials.

First, if anything should make subjects "careful enough" to search in a serial fashion, mixing class and orientation from trial to trial should. In fact, subjects had every opportunity to "be careful" in both of Krueger's experiments, since the display was terminated only after the subject responded. Second, if there is a need to "tune" some hypothetical feature extraction mechanism to be sensitive to those featural differences which would allow parallel search, it probably takes more than 24 trials. Thus, even in Krueger's Experiment 2, subjects probably had to adapt to a new target-background class and/or orientation difference before they had become sufficiently sensitive to the appropriate discriminating features for that block. In fact, given that Krueger's targets (5, 6, S, and G) were featurally similar to each other, it is plausible that subjects were searching for some feature(s) common to these targets - a nontrivial task if the orientation of the characters is varied as frequently as in Krueger (1984).

The second interesting aspect of the data is the significant difference in RT for the two between-category memory set size conditions in Experiment 2. I have argued above that the RT difference reflects different levels of encoding when M=1 as opposed to when M>1, as evidenced by elevated target present RT's when M>1 (Appendix A contains more precise formulations of this model and other models to follow). This RT difference has been found in at least three previous studies (Taylor, 1978, Hock et al., 1985, and Schneider and Shiffrin, 1977) using within-subjects designs, as well as in Experiments 1 and 2 of the present paper. All of these RT differences are of similar magnitude (Experiments 1 and 2, 72 and 81 ms, respectively; Hock et al., 49-51 ms, see Table 3, p. 78; Taylor, approximately 50 ms, see Figure 3, p. 433; Schneider and Shiffrin, approximately 50 ms, see Figure 5, p. 19) and are not terribly different in magnitude from the identification/categorization differences (82 ms, 106 ms, 87 ms, and 88 ms for Experiments 1-4, respectively) found by White (1977). I think that this RT difference, considered together with the catch trial data from Experiment 3, argue that search is based on different attributes of the stimuli in the two between-category, memory set size conditions.

The pattern in the two within-category conditions in the present experiments is quite different from that in the between-category conditions. Whatever strategy was being used in the M=1, within-category condition breaks down in the M=4, within-category condition (as evidenced by large slopes and high error rates). It seems reasonable to believe that subjects in these conditions were forced into some sort of serial comparison based on identity information (see Appendix A). On the other hand, the increase in memory set size made no difference (in slope) in the between-category conditions.

Alternative Hypotheses

I have suggested that this pattern of results indicates search based on featural information when M=1 and category information when M=4. It could be argued that the same featural information is being used in both between-category memory set size conditions. Cardosi (1986) has claimed that subjects set an overly lax criterion for responding "present" in between-category conditions, so that any gross physical difference (e.g., curvature, visual angle) is sufficient to allow correct detection in these conditions. Cardosi (1986, Experiments 1 and 2) found faster between-category response times for some targets, but not others. Since subjects were searching for four targets at once, Cardosi argued, this finding indicates that physical differences account for the category effect when M>1, as well as when M=1.

There are several problems with this interpretation. The subject's task in Cardosi's experiment was a forced-choice RT task, requiring one response if the target was one of two of the targets and the other response if it was one of the other two (e.g., "I" and "E" vs. "S" and "L"). After making this response, subjects told the experimenter which of the two targets (e.g., "I" or "E") within a group they had seen. As such, the task was one of identification (which should be sensitive to featural differences) with a target present on each trial. Thus, encoding and comparison on the basis of category membership would provide no advantage in this task as it presumably would in a standard two-choice ("present" or "absent") paradigm.

In fact, there is evidence indicating that it is necessary for subjects to be able to treat the memory set as an equivalence class in order for memory comparison to proceed in parallel, and that a task such as Cardosi's does not allow subjects to do so. Flach (1986) found no difference between M=2 and M=4 between-category, target-present conditions in a standard two-choice task, but did find longer RT's for M=4 in a "three-choice" task (there was no effect of display set size in either task condition). The three-choice task required subjects to respond with one button if no target was present, and one of two other buttons (according to which "subset" the target belonged to) if a target was present. In addition, Flach found that both the M=2 and M=4 conditions resulted in longer response times in the three-choice than the two choice task. Cardosi's subjects did not have the luxury of treating the memory set as an equivalence class, and so could not respond based on category membership.

Also, Cardosi did not vary display set size, making it impossible to know whether subjects were searching in parallel (an important diagnostic for the category effect). It is possible that subjects would have shown invariance of RT across display set size in the between- but not in the within-category conditions. If categorization is based on identification (as Dick (1971) argues), we might expect to find the same variation in RT's to specific targets which Cardosi argues rule out a category effect. Thus, convergent evidence from RT functions plotted against display set size would have been desirable.

Further, if subjects were responding based on the same (featural) information in both memory set size conditions (as Cardosi argued), one would not expect the means in these conditions to differ, since this information should be available at the same point in time. It could be argued, however, that gross featural differences allow easy localization in both memory set size conditions, after which the M=4 subjects perform a serial comparison of the featurally discrepant item with the memory set. Such a process would predict elevated target present RT's when M=4.

There are two sources of evidence against this alternative hypothesis. First, the catch trial data from Experiment 3 indicate that something quite different is going on in the two between-category memory set size conditions. If subjects were responding based on some gross featural difference between letters and digits when M=1, they should have false alarmed to the category foil - they did not. Additionally, if subjects were comparing the foil to the memory set when M=4, nine out of fourteen subjects should not have false alarmed to the category foil. The five M=4 subjects who did not false alarm to the category foil had a significantly higher slope than those who did. These five subjects also had a different pattern of errors than those subjects who did false alarm to the category foil (see Table 9). It could be that these subjects were attempting a serial search (given that their RT slopes are 49 ms/item and 23 ms/item for absent and present responses, respectively) based on featural/identity information, and that the required discriminations were too complex to be done in parallel as were discriminations based on category information. In any case, they were not the subjects "caught" by the featurally different (from the memory set) category foil.

Table 9

Error Rates for Two Groups of Subjects (those responding correctly vs. those responding incorrectly on the catch trial) in the M=4 Category Foil Condition of Experiment 3

	Response Incorrect (N=9)			Response Correct (N=5)		
	display set size			display set size		
Target	2	4	6	2	4	6
Present	.031	.030	.027	.012	.012	.060
Absent	.031	.037	.068	.008	.012	.028

Note. Since two of the subjects contributing data to this table were run in Experiment 2, means calculated from this table will not be equal to the corresponding values in Table 8.

However, one could entertain a particularly virulent version of the Cardosi (1986) argument which would ascribe between-category search performance to attention to item-specific features when M=1 in contrast to featural differences common to the category when M=4. If this was the case, it might account for the pattern of false alarms to the category foil. However, this alternative hypothesis does not hold up well for the particular stimuli in Experiment 3. Specifically, the numeral "3" appeared as a distractor during the regular trials, and as a distractor on the catch trial. As it turns out, the "3" has exactly the same height and width (and many of the same features) as the category foil "B". If gross featural differences were mediating performance when M=4, subjects should not have incorrectly responded "present" to something featurally similar to the distractor "3".

Second, there is evidence indicating that an extra check in an M>1, between-category condition would result in even higher RT's than those normally found. In the study by Jonides and Gleitman (1976) described on page 4, an extra between-category condition in which 25% of the between-category trials were "catch" trials (in this case, a member of the target set which was not a member of the memory set for that trial was present in the display) was added to the design. In order to be accurate, subjects would be forced to verify the identity of the digit in the display. This "modified between-category" condition resulted in a small slope, but elevated RT's (about 50 ms higher, see Figure 1, p. 292) at all levels of display set size (relative to that of the regular M=2 between-category condition). Jonides and Gleitman report that the zero-intercept values for these two conditions were significantly different. Although a more appropriate comparison might have been based on means or estimated N=1 intercepts, this result suggests that subjects do not normally perform a secondary check in between-category conditions when M>1.

There exists a second alternative hypothesis which cannot be ruled out so easily. On this hypothesis, subjects are using conceptual category information in both M=1 and M=4 between-category conditions, and categorization is based on identification. When M=1, it is assumed that presentation of the single target results in "priming"[3] (see Posner and Snyder, 1975 for a discussion of priming in similar paradigms) at the identity and category levels. When M=4, it is assumed (quite plausibly, given the catch trial data from Experiment 3) that subjects do not pay attention to the individual items in the memory set, so that only the category level is primed. Thus, presentation of the target should activate its identity-level code and hence its category code faster when M=1. Such a process would result in parallel RT functions in both conditions (since memory comparison is done using category information in both conditions), but a lower mean for the M=1 condition - exactly what was found in Experiment 2.

This alternative hypothesis also explains the catch trial data from Experiment 3 reasonably well. Subjects in the M=4, category foil condition should false alarm to "B", since they should get the same amount of "evidence" for target presence as they would get for any other letter. On the other hand, the amount (or rate of accumulation) of "activation" (or "evidence", if you prefer) generated by "B" (at the identity level, and hence at the category level) when M=1 would be lower than the corresponding activation for the actual memory set item ("H"). Thus, there is not enough evidence pointing to the presence of a target in the display to cause subjects in the M=1, category foil condition to make an erroneous "present" response. The feature foil data are also consistent with an account which assumes attention to category information and that categorization is based on identification (in both memory set size conditions). Let us assume that roughly the same proportion of subjects in both memory set size conditions identify the "4" as "H". The same subjects identifying the "4" as "H" will categorize the "4" as "letter", resulting in an erroneous "present" response. Recall that the M=1 and M=4 false alarm proportions were very similar (.57 and .50, respectively).

Explanation of the category effect by recourse to a priming process is not without precedent (cf., Deutsch, 1977). There is also empirical evidence (Gleitman and Jonides, 1978; Taylor, 1978, Experiment 4) indicating that subjects must develop a "set" for the target category in order to search in parallel in between-category conditions. Further, Posner and Snyder (1975) report evidence indicating that subjects must actively attend to a digit prime (analogous to the memory set in the present paradigm) in order to receive any benefit (i.e., decrease in RT) relative to a "neutral" prime ("+"). Posner and Snyder (1975, p. 69) asked subjects to indicate whether any digit was present in a row of letters which followed the prime. If subjects were instructed to deliberately match the prime to the array, they exhibited faster RT's than to the neutral prime; if they were not instructed to match the prime against the array, RT did not differ as a function of the type of prime (a digit or a "+"). When subjects were not instructed to attend specifically to the prime, they no doubt set themselves to search for "letter" (priming at the category level only, analogous to the M=4 condition in the present experiments). Unfortunately, Posner and Snyder (1975) did not report the magnitude of the difference between these two conditions.

The priming account would seem to be, on the surface, inconsistent with previous findings demonstrating that target-background similarity manipulations (as in Corcoran and Jackson, 1977) or feature-matching (as in Krueger, 1984) abolishes the category effect (since subjects are assumed to be utilizing category information in both M=1 and M>1 conditions). Remember, though, that category information is assumed to be derived from identity information. Also recall that "priming" contributes to the absolute level of RT, but that the type of information (identity or category) used for memory comparison determines whether the search will be parallel or serial. To the extent that increased similarity of target and background sets makes a parallel identification process unreliable (i.e., causes many false alarms due to crosstalk or causes hypothetical capacity limits to be exceeded), the categorization process will also be unreliable, and subjects might need to search in series in order to maintain acceptable accuracy.

For the same reasons discussed at the beginning of this chapter, Krueger's subjects may never have "discovered" or "tuned" a process based on category information. Even if they had discovered that they could search based on category information, such a strategy may not have worked in the next block. In Krueger's second experiment, the next block would have been a within-category block and/or the orientation of the characters might have changed enough to cause a new mapping of features to the identity level, possibly resulting in unreliable identification (if attempted in parallel).

There is, however, evidence which is not consistent with the priming hypothesis. Priming at the identity level should occur in both between- and within-category M=1 conditions. This implies that RT in the between-category, M=1 condition when there is a single element in the display (N=1) should be <u>longer</u> than the within-category, M=1, N=1 RT. This follows because the time to activate the appropriate identity code should be the same in both category conditions, and an extra step (which would take some time) would be required to categorize the display item in the between-category condition. Although N=1 data were not collected in the present experiments, the data points can be estimated. The estimated N=1 "intercepts" (496 ms, 537 ms, and 560 ms in the M=1, between-category condition, the M=1, within-category condition, and the M=4, between-category condition, respectively) are ordered as expected if subjects were searching at the featural, identity, and category level in the M=1, between-category, M=1, within-category, and M=4, between-category conditions, respectively. Further, the N=1 intercepts reported by Taylor (1978, Experiment 3) follow the same pattern (see Figure 4, p. 434) for his M=1, between-category, M=1, within-category, and "any member" between-category conditions.

Of course, these data are only suggestive, since they are projected from observed data to a point that was not actually measured. To illustrate, consider the case where the slope, if measured from N=1 to N=2, is much greater than the slope actually obtained from the data (resulting in a more curvilinear function, not very well described using linear parameters). Given this state of affairs, the N=1 intercept would be greater than the value which would be obtained if that point were actually measured. It is, however, a consistent difference, appearing in Experiment 1 and across different levels of practice in Taylor's third experiment. Still, these estimates are not conclusive evidence, and the question of whether the difference between the two between-category conditions is qualitative (reflecting search based on different stimulus attributes) or quantitative (reflecting different levels of priming) must await empirical data.

Speculations

It would appear then, that there is no simple framework which accounts for the "category effect" both when M=1 and when M>1. Based on his failure to replicate the "oh-zero" effect, Duncan (1983) argued that uncontrolled physical feature differences mediate the category effect when M=1. He then went on to speculate that, with increases in M, the category effect "...reflects the difference between well-learned and ill-learned classifications" (p. 231). I would like to expand on this distinction.

When M=1, we could view search based on identity as lying on a continuum. At one end (the "low" end, for present purposes) of the continuum are those target-distractor sets which are highly discriminable. In this case, it would take very little evidence (e.g., one or two target features) to decide that a particular display element is not the target. At the other end are those target-distractor sets which are not easily discriminable. In the most extreme cases, determining whether or not a display item is or is not the target would require an exhaustive comparison of its features with those of the target. At the "low" end, these comparisons could be expected to yield small slopes of RT against display set size, reflecting a parallel process. If we were to proceed "up" the continuum, we might find increasing slope values, until the slopes were in a 2:1 ratio, characteristic of a serial process. Whether this shift would be a qualitative one (from "parallel" to "serial" search) or a quantitative one (reflecting an "overloaded" parallel process) would be a point of some debate. Although slope values in a 2:1 ratio have usually been taken to indicate a serial process, Fisher (1986) has put forth a parallel model of search based on featural information which predicts increasing slope values with increasing target-distractor similarity. On this view, most between-category target-distractor sets would fall toward the "low" end of continuum, while most within-category sets would fall closer to the "high" end.

Although the slopes in the M=1, within-category conditions in Experiments 1 and 2 conformed to a roughly 2:1 ratio, they were not as large as those typically taken to indicate serial, self-terminating search based on identity information. Sternberg (1967a) has found a negative ("absent") slope of about 37 ms/item for both memory scanning and within-category visual search (with M=1). He also found positive ("present") slopes of about half the magnitude of the absent slopes in the visual search task. Thus, subjects may not be performing some sort of serial memory comparison with "abstract identity codes" when M=1, as was hypothesized initially. Instead, subjects may have been able to reject distractors rapidly (but still serially) based on a few features (see Appendix A). That is not to say that they would not have had to resort to comparison based on identity codes if the targets had been more (featurally) similar to the distractors. It is also possible, given the Fisher (1986) model, that the slopes in both M=1 conditions may reflect a parallel process (using featural information) which is less efficient in the M=1, within-category condition, and breaks down with increases in memory set size.

But what is happening when the memory set contains more than one item? In the present experiments, it is unlikely (in either category condition) that subjects would have been able to find a few physical attributes common to all four members of the memory set on which to base their search. This had dire consequences for subjects in the M=4, within-category conditions in both experiments. From subject reports, it seemed that the best strategy entailed rapid rehearsal of the memory set (one item at a time) in the hope that a display item would match an item being rehearsed before the display items were lost. Between-category subjects did not have to adopt this strategy - they simply encoded the display by category, comparing the "category codes" (in parallel) with the category ("letter") of the memory set.

This qualitative difference between the M=1 and M=4 between-category conditions in the present experiments is in harmony with the distinction between "display search" and "memory comparison" made by Flach (1986) and others. If featural information is available before category information, there is nothing to be gained by searching based on category information when M=1, since this strategy appears to result in longer response times (at least in the present experiment). However, when the memory load is increased (M=4) subjects are "induced" to treat the memory set as an equivalence class, allowing a single memory comparison based on category. Such a strategy is certainly preferable to a serial memory comparison (or rapid random matching) based on identity information, which is apparently what occurred in the M=4, within-category condition.

Methodological Considerations

The results of the present experiments make an important methodological point for future experimentation in this area. The fact that subjects may have been able to search in parallel in the M=1, between-category conditions of Experiments 2 and 3 (even though the featural differences between targets and distractors should have been roughly the same in the within-category conditions) emphasizes the importance of controlling featural differences. It is advisable to match carefully not only target sets and distractor sets on constituent features (e.g., as in Krueger, 1984), but target and distractor sets as well (i.e., control the absolute physical confusability of the target and distractor sets). There is some evidence that the category effect would still obtain under these conditions.

First, Dixon and Shedden (1987) found no category effect (in this case, higher accuracy in between-category conditions) with stimuli which were more confusable <u>between</u> than within categories (Experiment 1). However, they did find an effect when the stimuli were equally confusable both between and within categories (Experiment 2). It is difficult to compare their results with the present experiments, however, since they did not vary display set size within an experiment and used extremely brief (approximately 17 ms) exposure durations. Extrapolating to the present paradigm, the effect might be wiped out when M=1, (but not when M=4) by these more rigorous controls.

Second, Karlin and Bower (1976) found a between-category advantage in visual search for words with memory set sizes of three and six, but not when the memory set consisted of a single word. They argued that using words would make it impossible to attribute a between-category advantage to simple featural differences (since words are more visually complex than single characters). Their subjects reported performing a "featural" search based on global orthographic characteristics of the words when M=1 (Experiment 1), but not when M=3 (Experiment 2) or M=6 (Experiment 3). Karlin and Bower (1976) found almost identical slopes in a 2:1 ratio for both between- and within-category search when M=1 (Experiment 1), but significantly lower slopes for between-category search when M=3 and M=6. Between-category, M>1 RT slopes were far from flat (the lowest being 73 ms/item) and conformed to an approximately 2:1 ratio. However, this is not surprising, given that eye movements were clearly necessary, and the display was left on until the subject responded. It would appear that the same pattern of search across memory set sizes found in the present experiments obtains when simple featural differences cannot exist (since the stimuli were words).

Although an experiment holding all featural differences constant and varying memory set size would seem to be an appropriate follow-up to the present work, the idea has some limitations. How similar should the stimuli be? A "5" and an "S" composed of the same pixels are the same character. Even with more moderate degrees of similarity, a parallel process might become so susceptible to "crosstalk" (cf., Navon, 1986) and error-prone that it becomes impractical. Thus, in order to meet accuracy criteria, subjects might resort to a serial strategy. This would lead us to the erroneous conclusion that subjects never, under any conditions, search in parallel based on category membership.

Conclusions

I would like to close with a comment on the early/late selection question. It appears that drawing a strict dichotomy between early and late selection (at least with regard to present results) may be misguided. The idea that attention can operate at different levels of perceptual processing in response to different task demands is not new (cf., Johnston and Heinz, 1978). I have argued that selection in the between-category, M=1 conditions in the present experiments was based on physical feature differences. If that is correct, selection occurred relatively early in the course of perceptual processing. When the number of memory set items increased, discriminations based on featural information probably became too complex to be done in parallel, and subjects chose to base their search on another attribute of the stimulus (i.e., category membership). In this condition, selection occurred relatively late (after categorization of the stimuli) in the course of perceptual processing.

However, if the alternative hypothesis (discussed above) based on priming of identity or form when M=1 obtains, then both M=1 and M=4 between-category conditions respresent instances of selection based on conceptual category membership. As such, both conditions would represent instances of "late" selection. Regardless of which of these two hypotheses is correct, both the conceptual category effect and late selection (at least in this paradigm) are alive and well.

APPENDICES

APPENDIX A

SOME FORMULATIONS OF THE VARIOUS MODELS

Between-category:

Target-present conditions:

M=1: RT = e + d + r, where e is the time to encode an item in the display, d is the decision time (time required to make a parallel memory comparison based on features), and r is the time to initiate the appropriate motor response.

M=4: RT = e' + d + r, where e' is the time to encode an item in the display, d is the decision time (time required to make a parallel memory comparison based on category codes), and r is the time to initiate the appropriate motor response. e' = e + c, where c is the time to categorize an item in the display.

Target-absent conditions: For reasons described in Egeth et al. (1972), one would expect these slopes to increase given variable stimulus examination times (reflected in variation in d with N).

Within-category:

Target present and absent conditions:

M=1: If we assume serial, self-terminating search, then RT = $e + [(n+1)/2]d' + r$ when the target is present, and $e + nd' + r$ when the target is absent. e and r are as above, n is the display set size, and d' is the time required to make a single memory comparison based on item-specific information.

M=4: If we have serial, self-terminating search (display) and scanning (comparison) then RT = $e + \{[(m+1)/2][(n+1)/2]\}d'' + r$ when the target is present, and $e + mnd + r$ when the target is absent. e, r, and n are as above, m is the memory set size, and d'' is the time to make a single memory comparison based on item-specific information.

Notice that d' and d'' may or may not be equal. If both memory comparison processes operate on exactly the same information (e.g., "abstract" identity codes), then d' should equal d''. However, this predicts that the M=m slopes should be (m+1)/2 times larger than the M=1 slopes. When M=4, the slopes should be 5/2 = 2.5 times larger - they are closer to 3 times larger in Experiment 2 (although the M=4 slopes are quite variable). This suggests the possibility that, although serial, the M=1 comparison may have been made using one or two critical features to quickly reject each distractor. Such a process might result in a faster per item comparison time than one based on identity codes, particularly if the memory set items are stored "acoustically" when M=4 and have to be converted to some other form for the actual comparison process (as Sternberg (1967b) suggested).

APPENDIX B

ANOVA TABLES

Experiment 1 Reaction times

SOURCE	DF	SS	MS	F
SUBJ	7	994424.0000		
W1	1	1521368.0000	1521368.0000	21.1940
EW1B	7	502480.0000	71782.8600	
W2	1	1679072.0000	1679072.0000	75.6183
EW2B	7	155432.0000	22204.5700	
W3	1	732728.0000	732728.0000	84.6162
EW3B	7	60616.0000	8659.4290	
W4	2	576480.0000	288240.0000	33.6751
EW4B	14	119832.0000	8559.4290	
W12	1	610088.0000	610088.0000	77.6476
EW12B	7	55000.0000	7857.1430	
W13	1	7856.0000	7856.0000	1.8014
EW13B	7	30528.0000	4361.1430	
W14	2	118032.0000	59016.0000	10.0661
EW14B	14	82080.0000	5862.8570	
W23	1	21024.0000	21024.0000	5.6830
EW23B	7	25896.0000	3699.4280	
W24	2	186160.0000	93080.0000	10.9293
EW24B	14	119232.0000	8516.5710	
W34	2	164448.0000	82224.0000	20.0323
EW34B	14	57464.0000	4104.5710	
W123	1	19224.0000	19224.0000	3.5450
EW123B	7	37960.0000	5422.8570	
W124	2	73840.0000	36920.0000	4.8524
EW124B	14	106520.0000	7608.5710	
W134	2	28672.0000	14336.0000	9.5247
EW134B	14	21072.0000	1505.1430	
W234	2	29496.0000	14748.0000	5.5289
EW234B	14	37344.0000	2667.4280	
W1234	2	15328.0000	7664.0000	2.9335
EW1234B	14	36576.0000	2612.5720	
W	184	7231848.0000		

TSQ/N= 90067910.0000 N= 192 SST= 8226272.0000

WITHIN VARIABLE LABELS

1 = MEMORY SET SIZE

2 = CATEGORY (BETWEEN VS. WITHIN)

3 = TARGET PRESENCE/ABSENCE

4 = DISPLAY SET SIZE

Experiment 1 Slopes

SOURCE	DF	SS	MS	F
SUBJ	7	10450.9000		
W1	1	14455.2600	14455.2600	17.2036
EW1B	7	5881.7270	840.2466	
W2	1	23303.5600	23303.5600	16.1519
EW2B	7	10099.4500	1442.7780	
W3	1	19144.1800	19144.1800	22.8098
EW3B	7	5875.0860	839.2980	
W12	1	8790.9220	8790.9220	7.8324
EW12B	7	7856.6250	1122.3750	
W13	1	3501.3830	3501.3830	21.6544
EW13B	7	1131.8590	161.6942	
W23	1	3532.8050	3532.8050	10.4065
EW23B	7	2376.3670	339.4810	
W123	1	1947.6880	1947.6880	3.3753
EW123B	7	4039.3200	577.0458	
W	56	111936.2000		

TSQ/N= 70484.9400 N= 64 SST= 122387.1000

WITHIN VARIABLE LABELS

1 = MEMORY SET SIZE

2 = CATEGORY (BETWEEN VS. WITHIN)

3 = TARGET PRESENCE/ABSENCE

Experiment 2 Reaction times

SOURCE	DF	SS	MS	F
SUBJ	15	1323828.0000		
B1	1	399768.0000	399768.0000	6.0567
EB1	14	924060.0000	66004.2900	
W1	1	729268.0000	729268.0000	87.1675
W1B1	1	140020.0000	140020.0000	16.7362
EW1B1	14	117128.0000	8366.2860	
W2	1	387812.0000	387812.0000	61.9056
W2B1	1	32576.0000	32576.0000	5.2000
EW2B1	14	87704.0000	6264.5710	
W3	2	290932.0000	145466.0000	50.5040
W3B1	2	100944.0000	50472.0000	17.5233
EW3B1	28	80648.0000	2880.2860	
W12	1	8076.0000	8076.0000	2.1735
W12B1	1	29136.0000	29136.0000	7.8413
EW12B1	14	52020.0000	3715.7140	
W13	2	54900.0000	27450.0000	12.6091
W13B1	2	42736.0000	21368.0000	9.8153
EW13B1	28	60956.0000	2177.0000	
W23	2	63616.0000	31808.0000	22.8647
W23B1	2	18732.0000	9366.0000	6.7326
EW23B1	28	38952.0000	1391.1430	
W123	2	14932.0000	7466.0000	8.1468
W123B1	2	12088.0000	6044.0000	6.5952
EW123B1	28	25660.0000	916.4286	
W	176	2388836.0000		

TSQ/N= 54930590.0000 N= 192 SST= 3712664.0000

BETWEEN VARIABLE LABELS

1 = CATEGORY (BETWEEN VS. WITHIN)

WITHIN VARIABLE LABELS

1 = MEMORY SET SIZE

2 = TARGET PRESENCE/ABSENCE

3 = DISPSIZE

Experiment 2 Slopes

SOURCE	DF	SS	MS	F
SUBJ	15	20659.2600		
B1	1	12250.0600	12250.0600	20.3944
EB1	14	8409.1950	600.6568	
W1	1	6633.2930	6633.2930	15.2914
W1B1	1	5197.6800	5197.6800	11.9820
EW1B1	14	6073.0820	433.7916	
W2	1	7634.3910	7634.3910	27.2957
W2B1	1	2240.6020	2240.6020	8.0110
EW2B1	14	3915.6880	279.6920	
W12	1	1535.4570	1535.4570	9.0212
W12B1	1	1122.5860	1122.5860	6.5955
EW12B1	14	2382.8710	170.2051	
W	48	36735.6500		

TSQ/N= 35811.7700 N= 64 SST= 57394.9100

BETWEEN VARIABLE LABELS

1 = CATEGORY (BETWEEN VS. WITHIN)

WITHIN VARIABLE LABELS

1 = MEMORY SET SIZE

2 = TARGET PRESENCE/ABSENCE

Experiment 3 Reaction times

SOURCE	DF	SS	MS	F
SUBJ	47	4210552.0000		
B1	1	4904.0000	4904.0000	.0536
EB1	46	4205648.0000	91427.1300	
W1	1	364656.0000	364656.0000	62.4504
W1B1	1	2648.0000	2648.0000	.4535
EW1B1	46	268600.0000	5839.1300	
W2	2	233128.0000	116564.0000	60.9423
W2B1	2	19440.0000	9720.0000	5.0818
EW2B1	92	175968.0000	1912.6960	
W12	2	71840.0000	35920.0000	28.6047
W12B1	2	13992.0000	6996.0000	5.5712
EW12B1	92	115528.0000	1255.7390	
W	240	1265800.0000		

TSQ/N= 127968000.0000 N= 288 SST= 5476352.0000

BETWEEN VARIABLE LABELS

1 = MEMORY SET SIZE

WITHIN VARIABLE LABELS

1 = TARGET PRESENCE/ABSENCE

2 = DISPLAY SET SIZE

Experiment 3 Slopes

SOURCE	DF	SS	MS	F
SUBJ	47	17160.3700		
B1	1	2406.9060	2406.9060	7.5045
EB1	46	14753.4600	320.7274	
W1	1	8837.9590	8837.9590	39.3355
W1B1	1	1229.8670	1229.8670	5.4738
EW1B1	46	10335.3600	224.6816	
W	48	20403.1800		

TSQ/N= 27414.1700 N= 96 SST= 37563.5500

BETWEEN VARIABLE LABELS

1 = MEMORY SET SIZE

WITHIN VARIABLE LABELS

1 = TARGET PRESENCE/ABSENCE

NOTES

1. Experiments 2 and 3 were run at roughly the same time. In order to increase the number of observations in Experiment 3, between-category subjects in Experiment 2 were also given catch trials at the end of the second session. This resulted in an extra two subjects per cell in Experiment 3. It should be noted that the pattern of false alarms from Experiment 3 alone still deviates significantly from that expected by chance [$\chi^2(1) = 4.20$, $p < .05$; Fisher's exact probability = .017, two-tailed].

2. The same error criteria were applied to all conditions in Experiment 2 except the M=4, within-category condition (which proved to be an extremely difficult condition, even with the extended practice in Experiment 2). No subjects were replaced.

3. I am indebted to Alexander Pollatsek for this suggestion.

REFERENCES

Brand, J. (1971). Classification without identification in visual search. Quarterly Journal of Experimental Psychology, 23, 178-186.

Cardosi, K. M. (1986). Some determining factors of the alphanumeric category effect. Perception and Psychophysics, 40(5), 317-330.

Corcoran, D. W. J. and Jackson, A. (1977). Basic processes and strategies in visual search. In S. Dornic (Ed.), Attention and performance VI. Hillsdale, NJ: Erlbaum.

Deutsch, J. A. and Deutsch, D. (1963). Attention: Some theoretical considerations. Psychological Review, 70, 80-90.

Dick, A. O. (1971). Processing time for naming and categorization of letters and numbers. Perception and Psychophysics, 9(3B), 350-352.

Dixon, P. and Shedden, J. M. (1987). Conceptual and physical differences in the category effect. Perception and Psychophysics, 42(5), 457-464.

Duncan, J. (1983). Category effects in visual search: A failure to replicate the "oh-zero" phenomenon. Perception and Psychophysics, 34(3), 221-232.

Egeth, H. Jonides, J. and Wall, S. (1972). Parallel processing of multi-element displays. Cognitive Psychology, 3, 674-698.

Fisher, D. L. (1986). Hierarchical models of visual search: Serial and parallel processing. Paper presented at the annual meetings of the Society for Mathematical Psychology, Cambridge, MA.

Flach, J. M. (1986). Within-set discriminations in a consistent mapping search task. Perception and Psychophysics, 39(6), 397-406.

Gleitman, H. and Jonides, J. (1976). The cost of categorization in visual search: Incomplete processing of targets and field items. Perception and Psychophysics, 20, 281-288.

Gleitman, H. and Jonides, J. (1978). The effect of set on categorization in visual search. Perception and Psychophysics, 24, 361-368.

Hock, H., Rosenthal, A., and Stenquist, P. (1985). The category effect in visual search: Practice effects on catch trials. Perception and Psychophysics, 37(1), 73-80.

Ingling, N. W. (1972). Categorization: A mechanism for rapid information processing. Journal of Experimental Psychology, 94, 239-243.

Johnston, W. A. and Heinz, S. P. (1978). Flexibility and capacity demands of attention. Journal of Experimental Psychology: General, 107(4), 420-435.

Jonides, J. and Gleitman, H. (1972). A conceptual category effect in visual search: "O" as letter or digit. Perception and Psychophysics, 12, 457-460.

Jonides, J. and Gleitman, H. (1976). Benefit of categorization in visual search: Target location without identification. Perception and Psychophysics, 20, 289-298.

Karlin, M. B. and Bower, G. H. (1976). Semantic category effects in visual word search. Perception and Psychophysics, 19, 417-424.

Krueger, L. E. (1984). The category effect in visual search depends on physical rather than conceptual differences. Perception and Psychophysics, 35, 558-564.

Navon, D. (1985). Attention division or attention sharing? In M. I. Posner & O. S. M. Marin (Eds.), Attention and Performance XI. Hillsdale, NJ.: Erlbaum.

Navon, D. and Miller, J. (1987). Role of outcome conflict in dual-task performance. Journal of Experimental Psychology:Human Perception and Performance, 13(3), 435-448.

Nickerson, R. S. (1973). Can characters be classified directly as letters vs. digits or must they be identified first? Memory and Cognition, 1, 477-484.

Pollatsek, A., and Digman, L. (1977). Dependent spatial channels in visual processing. Cognitive Psychology, 9, 326-352.

Posner, M. (1970). On the relation between letter names and superordinate categories. Quarterly Journal of Experimental Psychology, 22, 279-287.

Posner, M. I., and Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), Information processing and cognition: The Loyola Symposium. Hillsdale, N.J.: Erlbaum.

Selfridge, O. G. (1959). Pandemonium: A paradigm for learning. In The Mechanisation of Thought Processes. London: H. M. Stationery Office.

Schneider, W. and Shiffrin, R. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. Psychological Review, 84(1), 1-66.

Shiffrin, R., Dumais, S., and Schneider, W. (1981). Characteristics of Automatism. In J. B. Long and A. D. Baddeley (Eds.), Attention and Performance IX. Hillsdale, NJ: Erlbaum.

Sperling, G., Budiansky, J., Spivak, J., and Johnson, M. (1971). Extremely rapid visual search: The maximum rate of scanning letters for the presence of a numeral. Science, 174, 307-311.

Sternberg, S. (1967a, April). Scanning a persisting visual image versus a memorized list. Paper presented at the annual meeting of the Eastern Psychological Association, Boston, MA.

Sternberg, S. (1967b). Two operations in character recognition: Some evidence from reaction-time measurements. Perception and Psychophysics, 2, 45-53.

Taylor, D. (1978). Identification and categorization of letters and digits. Journal of Experimental Psychology:Human Perception and Performance, 4, 423-439.

Townsend, J. T. (1971). A note on the identifiability of parallel and serial processes. Perception and Psychophysics, 10, 161-163.

Treisman, A. (1964). Verbal cues, language, and meaning in selective attention. American Journal of Psychology, 77, 206-218.

Treisman, A. and Gelade, H. (1980). A feature integration theory of attention. Cognitive Psychology, 12, 97-136.

Treisman, A. and Souther, J. (1985). Search asymmetry:A diagnostic for preattentive processing of separable features. Journal of Experimental Psychology:General, 114(3), 285-310.

White, M. (1977). Identification and categorization in visual search. Memory and Cognition, 5, 648-657.